2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2024)

Seattle, Washington, USA
16-22 June 2024

Pages 1-307

IEEE Catalog Number: CFP24003-POD
ISBN: 979-8-3503-5301-3

Copyright © 2024 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP24003-POD
ISBN (Print-On-Demand):	979-8-3503-5301-3
ISBN (Online):	979-8-3503-5300-6
ISSN:	1063-6919

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)

CVPR 2024

Table of Contents

Message from the Program and General Chairs ..... **cccxliii**
Organizers ..... **cccxlv**
Senior Area Chairs and Area Chairs ..... **cccxlviii**
Outstanding Reviewers ..... **ccclii**

Orals 1A Low-level Vision

Specularity Factorization for Low-Light Enhancement 1
Saurabh Saini (IIIT-Hyderabad, India) and P J Narayanan (IIIT-Hyderabad, India)

FlowIE: Efficient Image Enhancement via Rectified Flow 13
Yixuan Zhu (Tsinghua University, China), Wenliang Zhao (Tsinghua University, China), Ao Li (Tsinghua University, China), Yansong Tang (Tsinghua University, China), Jie Zhou (Tsinghua University, China), and Jiwen Lu (Tsinghua University, China)

Towards Robust Event-guided Low-Light Image Enhancement: A Large-Scale Real-World Event-Image Dataset and Novel Approach 23
Guoqiang Liang (Hong Kong University of Science and Technology(Guangzhou)), Kanghao Chen (Hong Kong University of Science and Technology(Guangzhou)), Hangyu Li (Hong Kong University of Science and Technology(Guangzhou)), Yunfan Lu (Hong Kong University of Science and Technology(Guangzhou)), and Lin Wang (Hong Kong University of Science and Technology(Guangzhou), Hong Kong University of Science and Technology)

Bilateral Event Mining and Complementary for Event Stream Super-Resolution 34
Zhilin Huang (Tsinghua University, China; Peng Cheng Laboratory, China), Quanmin Liang (Sun Yat-Sen University, China; Peng Cheng Laboratory, China), Yijie Yu (Tsinghua University, China; Peng Cheng Laboratory, China), Chujun Qin (China Southern Power Grid, China), Xiawu Zheng (Xiamen University, China; Peng Cheng Laboratory, China), Kai Huang (Sun Yat-Sen University, China), Zikun Zhou (Peng Cheng Laboratory, China), and Wenming Yang (Tsinghua University, China; Peng Cheng Laboratory, China)

FMA-Net: Flow-Guided Dynamic Filtering and Iterative Feature Refinement with Multi-Attention for Joint Video Super-Resolution and Deblurring 44
Geunhyuk Youk (KAIST), Jihyong Oh (Chung-Ang University), and Munchurl Kim (KAIST)

Orals 1B Vision and Graphics

GPLD3D: Latent Diffusion of 3D Shape Generative Models by Enforcing Geometric and Physical Priors 56
Yuan Dong (Institute for Intelligent Computing, Alibaba Group), Qi Zuo (Institute for Intelligent Computing, Alibaba Group), Xiaodong Gu (Institute for Intelligent Computing, Alibaba Group), Weihao Yuan (Institute for Intelligent Computing, Alibaba Group), Zhengyi Zhao (Institute for Intelligent Computing, Alibaba Group), Zilong Dong (Institute for Intelligent Computing, Alibaba Group), Liefeng Bo (Institute for Intelligent Computing, Alibaba Group), and Qixing Huang (The University of Texas at Austin)

Retrieval-Augmented Layout Transformer for Content-Aware Layout Generation 67
Daichi Horita (The University of Tokyo, Japan), Naoto Inoue (CyberAgent, Japan), Kotaro Kikuchi (CyberAgent, Japan), Kota Yamaguchi (CyberAgent, Japan), and Kiyoharu Aizawa (The University of Tokyo, Japan)

Eclipse: Disambiguating Illumination and Materials using Unintended Shadows 77
Dor Verbin (Google Research, USA), Ben Mildenhall (Google Research, USA), Peter Hedman (Google Research, UK), Jonathan T. Barron (Google Research, USA), Todd Zickler (Harvard University, USA), and Pratul P. Srinivasan (Google Research, USA)

Objects as Volumes: A Stochastic Geometry View of Opaque Solids 87
Bailey Miller (Carnegie Mellon University), Hanyu Chen (Carnegie Mellon University), Alice Lai (Carnegie Mellon University), and Ioannis Gkioulekas (Carnegie Mellon University)

DiffusionLight: Light Probes for Free by Painting a Chrome Ball 98
Pakkapon Phongthawee (VISTEC, Thailand), Worameth Chinchuthakun (Tokyo Institute of Technology, Japan), Nontaphat Sinsunthithet (VISTEC, Thailand), Varun Jampani (Stability AI, USA), Amit Raj (Google Research, USA), Pramook Khungurn (Pixiv, Japan), and Supasorn Suwajanakorn (VISTEC, Thailand)

Orals 1C Humans: Face, Body, Pose, Gesture, Movement

MultiPly: Reconstruction of Multiple People from Monocular Video in the Wild 109
Zeren Jiang (ETH Zurich, Switzerland), Chen Guo (ETH Zurich, Switzerland), Manuel Kaufmann (ETH Zurich, Switzerland), Tianjian Jiang (ETH Zurich, Switzerland), Julien Valentin (Microsoft, Switzerland), Otmar Hilliges (ETH Zurich, Switzerland), and Jie Song (ETH Zurich, Switzerland)

URHand: Universal Relightable Hands 119
Zhaoxi Chen (Nanyang Technological University), Gyeongsik Moon (Codec Avatars Lab, Meta), Kaiwen Guo (Codec Avatars Lab, Meta), Chen Cao (Codec Avatars Lab, Meta), Stanislav Pidhorskyi (Codec Avatars Lab, Meta), Tomas Simon (Codec Avatars Lab, Meta), Rohan Joshi (Codec Avatars Lab, Meta), Yuan Dong (Codec Avatars Lab, Meta), Yichen Xu (Codec Avatars Lab, Meta), Bernardo Pires (Codec Avatars Lab, Meta), He Wen (Codec Avatars Lab, Meta), Lucas Evans (Codec Avatars Lab, Meta), Bo Peng (Codec Avatars Lab, Meta), Julia Buffalini (Codec Avatars Lab, Meta), Autumn Trimble (Codec Avatars Lab, Meta), Kevyn McPhail (Codec Avatars Lab, Meta), Melissa Schoeller (Codec Avatars Lab, Meta), Shoou-I Yu (Codec Avatars Lab, Meta), Javier Romero (Codec Avatars Lab, Meta), Michael Zollhofer (Codec Avatars Lab, Meta), Yaser Sheikh (Codec Avatars Lab, Meta), Ziwei Liu (Nanyang Technological University), and Shunsuke Saito (Codec Avatars Lab, Meta)

Relightable Gaussian Codec Avatars 130
Shunsuke Saito (Codec Avatars Lab, Meta), Gabriel Schwartz (Codec Avatars Lab, Meta), Tomas Simon (Codec Avatars Lab, Meta), Junxuan Li (Codec Avatars Lab, Meta), and Giljoo Nam (Codec Avatars Lab, Meta)

Semantic Human Mesh Reconstruction with Textures 142
Xiaoyu Zhan (Nanjing University, China), Jianxin Yang (Nanjing University, China), Yuanqi Li (Nanjing University, China), Jie Guo (Nanjing University, China), Yanwen Guo (Nanjing University, China), and Wenping Wang (Texas A&M University, USA)

Stratified Avatar Generation from Sparse Observations 153
Han Feng (Wuhan University, China), Wenchao Ma (Pennsylvania State University, USA), Quankai Gao (University of Southern California, USA), Xianwei Zheng (Wuhan University, China), Nan Xue (Ant Group, China; Wuhan University, China), and Huijuan Xu (Pennsylvania State University, USA)

Poster Session 1

SEAS: ShapE-Aligned Supervision for Person Re-Identification 164
Haidong Zhu (University of Southern California), Pranav Budhwant (University of Southern California), Zhaoheng Zheng (University of Southern California), and Ram Nevatia (University of Southern California)

Test-Time Domain Generalization for Face Anti-Spoofing 175
Qianyu Zhou (Shanghai Jiao Tong University, China), Ke-Yue Zhang (Youtu Lab, Tencent, China), Taiping Yao (Youtu Lab, Tencent, China), Xuequan Lu (La Trobe University, Australia), Shouhong Ding (Youtu Lab, Tencent, China), and Lizhuang Ma (Shanghai Jiao Tong University, China)

Gradient Alignment for Cross-Domain Face Anti-Spoofing 188
Binh M. Le (Sungkyunkwan University) and Simon S. Woo (Sungkyunkwan University)

BigGait: Learning Gait Representation You Want by Large Vision Models 200
Dingqiang Ye (Southern University of Science and Technology, China), Chao Fan (Southern University of Science and Technology, China), Jingzhe Ma (Southern University of Science and Technology, China), Xiaoming Liu (Michigan State University, USA), and Shiqi Yu (Southern University of Science and Technology, China)

Suppress and Rebalance: Towards Generalized Multi-Modal Face Anti-Spoofing 211
Xun Lin (Beihang University), Shuai Wang (Beihang University), Rizhao Cai (Nanyang Technological University), Yizhong Liu (Beihang University), Ying Fu (Beijing Institute of Technology), Wenzhong Tang (Beihang University), Zitong Yu (Great Bay University), and Alex Kot (Nanyang Technological University)

CFPL-FAS: Class Free Prompt Learning for Generalizable Face Anti-spoofing 222
Ajian Liu (Chinese Academy of Sciences, China), Shuai Xue (Beijing Institute of Technology, China), Jianwen Gan (Wuzhou University, China), Jun Wan (Chinese Academy of Sciences, China), Yanyan Liang (Macau University of Science and Technology, China), Jiankang Deng (Imperial College London, UK), Sergio Escalera (Universitat de Barcelona and Computer Vision Center, Spain), and Zhen Lei (Chinese Academy of Sciences, China)

Psychometry: An Omnifit Model for Image Reconstruction from Human Brain Activity 233
Ruijie Quan (Zhejiang University), Wenguan Wang (Zhejiang University), Zhibo Tian (Lanzhou University), Fan Ma (Zhejiang University), and Yi Yang (Zhejiang University)

KeyPoint Relative Position Encoding for Face Recognition 244
Minchul Kim (Michigan State University), Yiyang Su (Michigan State University), Feng Liu (Michigan State University), Anil Jain (Michigan State University), and Xiaoming Liu (Michigan State University)

Distilling CLIP with Dual Guidance for Learning Discriminative Human Body Shape Representation 256
Feng Liu (Michigan State University), Minchul Kim (Michigan State University), Zhiyuan Ren (Michigan State University), and Xiaoming Liu (Michigan State University)

Flexible Biometrics Recognition: Bridging the Multimodality Gap through Attention, Alignment and Prompt Tuning 267
Leslie Ching Ow Tiong (Samsung Electronics, South Korea), Dick Sigmund (AIDOT Inc., South Korea), Chen-Hui Chan (Korea Institute of Science and Technology, South Korea), and Andrew Beng Jin Teoh (Yonsei University, South Korea)

One-Class Face Anti-spoofing via Spoof Cue Map-Guided Feature Learning 277
Pei-Kai Huang (National Tsing Hua University), Cheng-Hsuan Chiang (National Tsing Hua University), Tzu-Hsien Chen (National Tsing Hua University), Jun-Xiong Chong (National Tsing Hua University), Tyng-Luh Liu (Academia Sinica), and Chiou-Ting Hsu (National Tsing Hua University)

Activity-Biometrics: Person Identification from Daily Activities 287
Shehreen Azad (University of Central Florida, USA) and Yogesh Singh Rawat (University of Central Florida, USA)

Privacy-Preserving Face Recognition Using Trainable Feature Subtraction 297
Yuxi Mi (Fudan University), Zhizhou Zhong (Fudan University), Yuge Huang (Youtu Lab, Tencent), Jiazhen Ji (Youtu Lab, Tencent), Jianqing Xu (Youtu Lab, Tencent), Jun Wang (WeChat Pay Lab33, Tencent), Shaoming Wang (WeChat Pay Lab33, Tencent), Shouhong Ding (Youtu Lab, Tencent), and Shuigeng Zhou (Fudan University)

Molecular Data Programming: Towards Molecule Pseudo-labeling with Systematic Weak Supervision 308
Xin Juan (Jilin University, China), Kaixiong Zhou (Massachusetts Institute of Technology, USA), Ninghao Liu (University of Georgia, USA), Tianlong Chen (Massachusetts Institute of Technology, USA), and Xin Wang (Jilin University, China)

Clustering for Protein Representation Learning 319
Ruijie Quan (Zhejiang University), Wenguan Wang (Zhejiang University), Fan Ma (Zhejiang University), Hehe Fan (Zhejiang University), and Yi Yang (Zhejiang University)

Fun with Flags: Robust Principal Directions via Flag Manifolds 330
Nathan Mankovich (University of Valencia), Gustau Camps-Valls (University of Valencia), and Tolga Birdal (Imperial College)

CAM Back Again: Large Kernel CNNs from a Weakly Supervised Object Localization Perspective. 341
Shunsuke Yasuki (Rikkyo University) and Masato Taki (Rikkyo University)

Confronting Ambiguity in 6D Object Pose Estimation via Score-Based Diffusion on SE(3) 352
Tsu-Ching Hsiao (National Tsing Hua University, Taiwan), Hao-Wei Chen (National Tsing Hua University, Taiwan), Hsuan-Kung Yang (National Tsing Hua University, Taiwan), and Chun-Yi Lee (National Tsing Hua University, Taiwan)

Quantifying Task Priority for Multi-Task Optimization 363
Wooseong Jeong (KAIST) and Kuk-Jin Yoon (KAIST)

Unbiased Estimator for Distorted Conics in Camera Calibration 373
Chaehyeon Song (Seoul National University), Jaeho Shin (Seoul National University), Myung-Hwan Jeon (Seoul National University), Jongwoo Lim (Seoul National University), and Ayoung Kim (Seoul National University)

Multi-Object Tracking in the Dark 382
Xinzhe Wang (Beijing Institute of Technology, China), Kang Ma (Beijing Institute of Technology, China), Qiankun Liu (Beijing Institute of Technology, China), Yunhao Zou (Beijing Institute of Technology, China), and Ying Fu (Beijing Institute of Technology, China)

Implicit Discriminative Knowledge Learning for Visible-Infrared Person Re-Identification 393
Kaijie Ren (Chongqing University, China) and Lei Zhang (Chongqing University, China)

From Correspondences to Pose: Non-minimal Certifiably Optimal Relative Pose without Disambiguation 403
Javier Tirado-Garín (I3A, University of Zaragoza) and Javier Civera (I3A, University of Zaragoza)

From Activation to Initialization: Scaling Insights for Optimizing Neural Fields 413
Hemanth Saratchandran (Australian Institute of Machine Learning, University of Adelaide), Sameera Ramasinghe (Amazon), and Simon Lucey (Australian Institute of Machine Learning, University of Adelaide)

PairDETR : Joint Detection and Association of Human Bodies and Faces 423
Ammar Ali (MTS AI, ITMO), Georgii Gaikov (MTS AI), Denis Rybalchenko (VisionLabs), Alexander Chigorin (VisionLabs), Ivan Laptev (MBZUAI), and Sergey Zagoruyko (MTS AI, Skoltech)

Move as You Say, Interact as You Can: Language-guided Human Motion Generation with Scene Affordance 433
Zan Wang (Beijing Institute of Technology, China; Beijing Institute for General Artificial Intelligence, China), Yixin Chen (Beijing Institute for General Artificial Intelligence, China), Baoxiong Jia (Beijing Institute for General Artificial Intelligence, China), Puhao Li (Tsinghua University, China; Beijing Institute for General Artificial Intelligence, China), Jinlu Zhang (Peking University, China; Beijing Institute for General Artificial Intelligence, China), Jingze Zhang (Tsinghua University, China; Beijing Institute for General Artificial Intelligence, China), Tengyu Liu (Beijing Institute for General Artificial Intelligence, China), Yixin Zhu (Peking University, China), Wei Liang (Beijing Institute of Technology, China; Yangtze Delta Region Academy of Beijing Institute of Technology, Jiaxing, China), and Siyuan Huang (Beijing Institute for General Artificial Intelligence, China)

OAKINK2: A Dataset of Bimanual Hands-Object Manipulation in Complex Task Completion 445
Xinyu Zhan (Shanghai Jiao Tong University, China), Lixin Yang (Shanghai Jiao Tong University, China), Yifei Zhao (Shanghai Jiao Tong University, China), Kangrui Mao (Shanghai Jiao Tong University, China), Hanlin Xu (Shanghai Jiao Tong University, China), Zenan Lin (South China University of Technology, China), Kailin Li (Shanghai Jiao Tong University, China), and Cewu Lu (Shanghai Jiao Tong University, China)

Seamless Human Motion Composition with Blended Positional Encodings 457
German Barquero (Universitat de Barcelona and Computer Vision Center, Spain), Sergio Escalera (Universitat de Barcelona and Computer Vision Center, Spain), and Cristina Palmero (Universitat de Barcelona and Computer Vision Center, Spain)

VideoRF: Rendering Dynamic Radiance Fields as 2D Feature Video Streams 470
Liao Wang (ShanghaiTech University, China), Kaixin Yao (ShanghaiTech University, China), Chengcheng Guo (ShanghaiTech University, China), Zhirui Zhang (ShanghaiTech University, China), Qiang Hu (Shanghai Jiao Tong University, China), Jingyi Yu (ShanghaiTech University, China), Lan Xu (ShanghaiTech University, China), and Minye Wu (KU Leuven, Belgium)

OMG: Towards Open-vocabulary Motion Generation via Mixture of Controllers 482
Han Liang (ShanghaiTech University, China), Jiacheng Bao (ShanghaiTech University, China), Ruichi Zhang (ShanghaiTech University, China), Sihan Ren (ShanghaiTech University, China), Yuecheng Xu (ShanghaiTech University, China), Sibei Yang (ShanghaiTech University, China), Xin Chen (Tencent PCG, China), Jingyi Yu (ShanghaiTech University, China), and Lan Xu (ShanghaiTech University, China)

HOLD: Category-agnostic 3D Reconstruction of Interacting Hands and Objects from Video 494
Zicong Fan (ETH Zürich, Switzerland; Max Planck Institute for Intelligent Systems, Tübingen, Germany), Maria Parelli (ETH Zürich, Switzerland), Maria Eleni Kadoglou (ETH Zürich, Switzerland), Xu Chen (ETH Zürich, Switzerland; Max Planck Institute for Intelligent Systems, Tübingen, Germany), Muhammed Kocabas (ETH Zürich, Switzerland; Max Planck Institute for Intelligent Systems, Tübingen, Germany), Michael J. Black (Max Planck Institute for Intelligent Systems, Tübingen, Germany), and Otmar Hilliges (ETH Zürich, Switzerland)

HUGS: Human Gaussian Splats 505
Muhammed Kocabas (Apple, USA), Jen-Hao Rick Chang (Apple, USA), James Gabriel (Apple, USA), Oncel Tuzel (Apple, USA), and Anurag Ranjan (Apple, USA)

HOI-M^3: Capture Multiple Humans and Objects Interaction within Contextual Environment 516
Juze Zhang (ShanghaiTech University, China;University of Chinese Academy of Sciences, China), Jingyan Zhang (ShanghaiTech University, China), Zining Song (ShanghaiTech University, China), Zhanhe Shi (ShanghaiTech University, China), Chengfeng Zhao (ShanghaiTech University, China), Ye Shi (ShanghaiTech University, China), Jingyi Yu (ShanghaiTech University, China), Lan Xu (ShanghaiTech University, China), and Jingya Wang (ShanghaiTech University, China)

InterHandGen: Two-Hand Interaction Generation via Cascaded Reverse Diffusion 527
Jihyun Lee (KAIST), Shunsuke Saito (Meta), Giljoo Nam (Meta), Minhyuk Sung (KAIST), and Tae-Kyun Kim (KAIST, Imperial College London)

SiTH: Single-view Textured Human Reconstruction with Image-Conditioned Diffusion 538
Hsuan- I Ho (ETH Zürich, Switzerland), Jie Song (ETH Zürich, Switzerland), and Otmar Hilliges (ETH Zürich, Switzerland)

4D-DRESS: A 4D Dataset of Real-World Human Clothing With Semantic Annotations 550
Wenbo Wang (ETH Zürich, Switzerland), Hsuan-I Ho (ETH Zürich, Switzerland), Chen Guo (ETH Zürich, Switzerland), Boxiang Rong (ETH Zürich, Switzerland), Artur Grigorev (ETH Zürich, Switzerland), Jie Song (ETH Zürich, Switzerland), Juan Jose Zarate (ETH Zürich, Switzerland), and Otmar Hilliges (ETH Zürich, Switzerland)

FinePOSE: Fine-Grained Prompt-Driven 3D Human Pose Estimation via Diffusion Models 561
Jinglin Xu (University of Science and Technology Beijing, China), Yijie Guo (Peking University, China), and Yuxin Peng (Peking University, China)

Real-Time Simulated Avatar from Head-Mounted Sensors 571
Zhengyi Luo (Carnegie Mellon University, USA), Jinkun Cao (Carnegie Mellon University, USA), Rawal Khirodkar (Meta Platforms, Inc., USA), Alexander Winkler (Meta Platforms, Inc., USA), Kris Kitani (Carnegie Mellon University, USA), and Weipeng Xu (Meta Platforms, Inc., USA)

Digital Life Project: Autonomous 3D Characters with Social Intelligence 582
Zhongang Cai (S-Lab, Nanyang Technological University, Singapore), Jianping Jiang (SenseTime Research, China), Zhongfei Qing (SenseTime Research, China), Xinying Guo (S-Lab, Nanyang Technological University, Singapore), Mingyuan Zhang (S-Lab, Nanyang Technological University, Singapore), Zhengyu Lin (SenseTime Research, China), Haiyi Mei (SenseTime Research, China), Chen Wei (SenseTime Research, Singapore), Ruisi Wang (S-Lab, Nanyang Technological University, Singapore), Wanqi Yin (SenseTime Research, Singapore), Liang Pan (S-Lab, Nanyang Technological University, Singapore), Xiangyu Fan (SenseTime Research, China), Han Du (SenseTime Research, China), Peng Gao (SenseTime Research, China), Zhitao Yang (SenseTime Research, China), Yang Gao (SenseTime Research, China), Jiaqi Li (SenseTime Research, China), Tianxiang Ren (SenseTime Research, China), Yukun Wei (SenseTime Research, China), Xiaogang Wang (SenseTime Research, China), Chen Change Loy (S-Lab, Nanyang Technological University, Singapore), Lei Yang (SenseTime Research, China), and Ziwei Liu (S-Lab, Nanyang Technological University, Singapore)

Learning Visual Prompt for Gait Recognition 593
Kang Ma (Beijing Institute of Technology, China), Ying Fu (Beijing Institute of Technology, China), Chunshui Cao (WATRIX.AI, China), Saihui Hou (Beijing Normal University, China), Yongzhen Huang (Beijing Normal University, China), and Dezhi Zheng (Beijing Institute of Technology, China)

Hourglass Tokenizer for Efficient Transformer-Based 3D Human Pose Estimation 604
Wenhao Li (National Key Laboratory of General Artificial Intelligence, Peking University, Shenzhen Graduate School), Mengyuan Liu (National Key Laboratory of General Artificial Intelligence, Peking University, Shenzhen Graduate School), Hong Liu (National Key Laboratory of General Artificial Intelligence, Peking University, Shenzhen Graduate School), Pichao Wang (Amazon Prime Video), Jialun Cai (National Key Laboratory of General Artificial Intelligence, Peking University, Shenzhen Graduate School), and Nicu Sebe (University of Trento)

LocLLM: Exploiting Generalizable Human Keypoint Localization via Large Language Model 614
Dongkai Wang (Peking University, China), Shiyu Xuan (Peking University, China), and Shiliang Zhang (Peking University, China)

Spatial-Aware Regression for Keypoint Localization 624
Dongkai Wang (Peking University, China) and Shiliang Zhang (Peking University, China)

GaussianAvatar: Towards Realistic Human Avatar Modeling from a Single Video via Animatable 3D Gaussians 634
Liangxiao Hu (Harbin Institute of Technology, China), Hongwen Zhang (Beijing Normal University, China), Yuxiang Zhang (Tsinghua University, China), Boyao Zhou (Tsinghua University, China), Boning Liu (Tsinghua University, China), Shengping Zhang (Harbin Institute of Technology, China), and Liqiang Nie (Harbin Institute of Technology, China)

HHMR: Holistic Hand Mesh Recovery by Enhancing the Multimodal Controllability of Graph Diffusion Models 645
Mengcheng Li (Tsinghua University, China), Hongwen Zhang (Beijing Normal University, China), Yuxiang Zhang (Tsinghua University, China), Ruizhi Shao (Tsinghua University, China), Tao Yu (Tsinghua University, China), and Yebin Liu (Tsinghua University, China)

Capturing Closely Interacted Two-Person Motions with Reaction Priors 655
Qi Fang (NetEase Games AI Lab, China), Yinghui Fan (NetEase Games AI Lab, China), Yanjun Li (NetEase Games AI Lab, China), Junting Dong (Shanghai AI Lab, China), Dingwei Wu (NetEase Games AI Lab, China), Weidong Zhang (NetEase Games AI Lab, China), and Kang Chen (NetEase Games AI Lab, China)

SyncTalk: The Devil is in the Synchronization for Talking Head Synthesis 666
Ziqiao Peng (Renmin University of China), Wentao Hu (Beijing University of Posts and Telecommunications), Yue Shi (Psyche AI Inc.), Xiangyu Zhu (Chinese Academy of Sciences), Xiaomei Zhang (Chinese Academy of Sciences), Hao Zhao (Tsinghua University), Jun He (Renmin University of China), Hongyan Liu (Tsinghua University), and Zhaoxin Fan (Renmin University of China)

Single-to-Dual-View Adaptation for Egocentric 3D Hand Pose Estimation 677
Ruicong Liu (The University of Tokyo, Japan), Takehiko Ohkawa (The University of Tokyo, Japan), Mingfang Zhang (The University of Tokyo, Japan), and Yoichi Sato (The University of Tokyo, Japan)

Bidirectional Autoregessive Diffusion Model for Dance Generation 687
Canyu Zhang (University of South Carolina, USA), Youbao Tang (PAII Inc., USA), Ning Zhang (PAII Inc., USA), Ruei-Sung Lin (PAII Inc., USA), Mei Han (PAII Inc., USA), Jing Xiao (Ping An Technology, China), and Song Wang (University of South Carolina, USA)

High-Quality Facial Geometry and Appearance Capture at Home 697
Yuxuan Han (Tsinghua University, China), Junfeng Lyu (Tsinghua University, China), and Feng Xu (Tsinghua University, China)

Multiple View Geometry Transformers for 3D Human Pose Estimation 708
Ziwei Liao (University of Toronto, Canada), Jialiang Zhu (Southeast University, China), Chunyu Wang (Microsoft Research Asia, China), Han Hu (Microsoft Research Asia, China), and Steven L. Waslander (University of Toronto, Canada)

PACER+: On-Demand Pedestrian Animation Controller in Driving Scenarios 718
Jingbo Wang (Shanghai AI LAB, China), Zhengyi Luo (Carnegie Mellon University, USA), Ye Yuan (NVIDIA, USA), Yixuan Li (The Chinese University of Hong Kong, Hong Kong), and Bo Dai (Shanghai AI LAb, China)

I'M HOI: Inertia-aware Monocular Capture of 3D Human-Object Interactions 729
Chengfeng Zhao (ShanghaiTech University, China), Juze Zhang (ShanghaiTech University, China), Jiashen Du (ShanghaiTech University, China), Ziwei Shan (ShanghaiTech University, China), Junye Wang (ShanghaiTech University, China), Jingyi Yu (ShanghaiTech University, China), Jingya Wang (ShanghaiTech University, China), and Lan Xu (ShanghaiTech University, China)

HAVE-FUN: Human Avatar Reconstruction from Few-Shot Unconstrained Images 742
Xihe Yang (The Chinese University of Hong Kong (Shenzhen), China), Xingyu Chen (Xiaobing.AI), Daiheng Gao (Freelancer), Shaohui Wang (Tsinghua University, China), Xiaoguang Han (The Chinese University of Hong Kong (Shenzhen), China), and Baoyuan Wang (Xiaobing.AI)

Can Language Beat Numerical Regression? Language-Based Multimodal Trajectory Prediction 753
Inhwan Bae (Gwangju Institute of Science and Technology, South Korea), Junoh Lee (Gwangju Institute of Science and Technology, South Korea), and Hae-Gon Jeon (Gwangju Institute of Science and Technology, South Korea)

3D Human Pose Perception from Egocentric Stereo Videos 767
Hiroyasu Akada (Max Planck Institute for Informatics, SIC), Jian Wang (Max Planck Institute for Informatics, SIC), Vladislav Golyanik (Max Planck Institute for Informatics, SIC), and Christian Theobalt (Max Planck Institute for Informatics, SIC)

Egocentric Whole-Body Motion Capture with FisheyeViT and Diffusion-Based Motion Refinement..... 777
Jian Wang (Max Planck Institute for Informatics, Germany), Zhe Cao (Google, USA), Diogo Luvizon (Max-Planck Institute for Informatics, Germany), Lingjie Liu (University of Pennsylvania, USA), Kripasindhu Sarkar (Google, Switzerland), Danhang Tang (Google, USA), Thabo Beeler (Google, Switzerland), and Christian Theobalt (Max Planck Institute for Informatics, Germany)

Human Gaussian Splatting: Real-time Rendering of Animatable Avatars 788
Arthur Moreau (Huawei Noah's Ark Lab), Jifei Song (Huawei Noah's Ark Lab), Helisa Dhamo (Huawei Noah's Ark Lab), Richard Shaw (Huawei Noah's Ark Lab), Yiren Zhou (Huawei Noah's Ark Lab), and Eduardo Pérez-Pellitero (Huawei Noah's Ark Lab)

OHTA: One-shot Hand Avatar via Data-driven Implicit Priors 799
Xiaozheng Zheng (ByteDance, China), Chao Wen (ByteDance, China), Zhuo Su (ByteDance, China), Zeran Xu (ByteDance, China), Zhaohu Li (ByteDance, China), Yang Zhao (ByteDance, China), and Zhou Xue (ByteDance, China)

HOIAnimator: Generating Text-prompt Human-object Animations using Novel Perceptive Diffusion Models 811
Wenfeng Song (Beijing Information Science and Technology University, China), Xinyu Zhang (Beijing Information Science and Technology University), Shuai Li (Zhongguancun Laboratory, China), Yang Gao (State Key Laboratory of Virtual Reality Technology and Systems, Beihang University), Aimin Hao (State Key Laboratory of Virtual Reality Technology and Systems, Beihang University), Xia Hou (Beijing Information Science and Technology University), Chenglizhao Chen (China University of Petroleum (East China)), Ning Li (Beijing Information Science and Technology University), and Hong Qin (Stony Brook University (SUNY at Stony Brook), New York)

Arbitrary Motion Style Transfer with Multi-condition Motion Latent Diffusion Model 821
Wenfeng Song (Beijing Information Science and Technology University, China), Xingliang Jin (Beijing Information Science and Technology University, China), Shuai Li (Zhongguancun Laboratory, China), Chenglizhao Chen (China University of Petroleum (East China)), Aimin Hao (State Key Laboratory of Virtual Reality Technology and Systems, Beihang University), Xia Hou (Beijing Information Science and Technology University, China), Ning Li (Beijing Information Science and Technology University, China), and Hong Qin (Stony Brook University (SUNY at Stony Brook), New York, USA)

Single-View Scene Point Cloud Human Grasp Generation 831
Yan-Kang Wang (Sun Yat-sen University, China), Chengyi Xing (Stanford University, Stanford, CA), Yi-Lin Wei (Sun Yat-sen University, China), Xiao-Ming Wu (Sun Yat-sen University, China), and Wei-Shi Zheng (Sun Yat-sen University, China)

Attention-Propagation Network for Egocentric Heatmap to 3D Pose Lifting 842
Taeho Kang (Seoul National University) and Youngki Lee (Seoul National University)

AnySkill: Learning Open-Vocabulary Physical Skill for Interactive Agents 852
Jieming Cui (Peking University), Tengyu Liu (National Key Laboratory of General Artificial Intelligence, BIGAI), Nian Liu (Beijing University of Posts and Telecommunications), Yaodong Yang (Peking University), Yixin Zhu (Peking University), and Siyuan Huang (National Key Laboratory of General Artificial Intelligence, BIGAI)

From a Bird's Eye View to See: Joint Camera and Subject Registration without the Camera Calibration 863
Zekun Qian (Tianjin University, China), Ruize Han (Chinese Academy of Sciences, China), Wei Feng (Tianjin University, China), and Song Wang (University of South Carolina, USA)

HMD-Poser: On-Device Real-time Human Motion Tracking from Scalable Sparse Observations 874
Peng Dai (PICO, ByteDance), Yang Zhang (PICO, ByteDance), Tao Liu (PICO, ByteDance), Zhen Fan (PICO, ByteDance), Tianyuan Du (PICO, ByteDance), Zhuo Su (PICO, ByteDance), Xiaozheng Zheng (PICO, ByteDance), and Zeming Li (PICO, ByteDance)

Monocular Identity-Conditioned Facial Reflectance Reconstruction 885
Xingyu Ren (Shanghai Jiao Tong University, China), Jiankang Deng (Imperial College London, UK), Yuhao Cheng (Shanghai Jiao Tong University, China), Jia Guo (InsigtFace, China), Chao Ma (Shanghai Jiao Tong University, China), Yichao Yan (Shanghai Jiao Tong University, China), Wenhan Zhu (Shanghai Jiao Tong University, China), and Xiaokang Yang (Shanghai Jiao Tong University, China)

GAvatar: Animatable 3D Gaussian Avatars with Implicit Mesh Learning 896
Ye Yuan (NVIDIA), Xueting Li (NVIDIA), Yangyi Huang (NVIDIA), Shalini De Mello (NVIDIA), Koki Nagano (NVIDIA), Jan Kautz (NVIDIA), and Umar Iqbal (NVIDIA)

Score-Guided Diffusion for 3D Human Recovery 906
Anastasis Stathopoulos (Rutgers University), Ligong Han (Rutgers University), and Dimitris Metaxas (Rutgers University)

3D-Aware Face Editing via Warping-Guided Latent Direction Learning 916
Yuhao Cheng (Shanghai Jiao Tong University, China), Zhuo Chen (Shanghai Jiao Tong University, China), Xingyu Ren (Shanghai Jiao Tong University, China), Wenhan Zhu (Shanghai Jiao Tong University, China), Zhengqin Xu (Shanghai Jiao Tong University, China), Di Xu (Huawei Cloud Computing Technologies Co., Ltd, China), Changpeng Yang (Huawei Cloud Computing Technologies Co., Ltd, China), and Yichao Yan (Shanghai Jiao Tong University, China)

WANDR: Intention-guided Human Motion Generation 927
Markos Diomataris (Max Planck Institute for Intelligent Systems), Nikos Athanasiou (Max Planck Institute for Intelligent Systems), Omid Taheri (Max Planck Institute for Intelligent Systems), Xi Wang (ETH Zurich), Otmar Hilliges (ETH Zurich), and Michael J. Black (Max Planck Institute for Intelligent Systems)

Exploring Vision Transformers for 3D Human Motion-Language Models with Motion Patches 937
Qing Yu (LY Corporation, Japan), Mikihiro Tanaka (LY Corporation, Japan), and Kent Fujiwara (LY Corporation, Japan)

NIFTY: Neural Object Interaction Fields for Guided Human Motion Synthesis 947
Nilesh Kulkarni (University of Michigan, USA), Davis Rempe (NVIDIA, USA), Kyle Genova (Google, USA), Abhijit Kundu (Google, USA), Justin Johnson (University of Michigan, USA), David Fouhey (New York University, USA), and Leonidas Guibas (Stanford University, USA)

DreamAvatar: Text-and-Shape Guided 3D Human Avatar Generation via Diffusion Models 958
Yukang Cao (The University of Hong Kong), Yan-Pei Cao (Tencent), Kai Han (The University of Hong Kong), Ying Shan (Tencent), and Kwan-Yee K. Wong (The University of Hong Kong)

Person-in-WiFi 3D: End-to-End Multi-Person 3D Pose Estimation with Wi-Fi 969
Kangwei Yan (Xi'an Jiaotong University), Fei Wang (Xi'an Jiaotong University), Bo Qian (Xi'an Jiaotong University), Han Ding (Xi'an Jiaotong University), Jinsong Han (Zhejiang University), and Xing Wei (Xi'an Jiaotong University)

ScoreHypo: Probabilistic Human Mesh Estimation with Hypothesis Scoring 979
Yuan Xu (Peking University, China), Xiaoxuan Ma (Peking University, China), Jiajun Su (International Digital Economy Academy (IDEA), China), Wentao Zhu (Peking University, China), Yu Qiao (Shanghai Jiao Tong University, China), and Yizhou Wang (Peking University, China)

Relightable and Animatable Neural Avatar from Sparse-View Video 990
Zhen Xu (Zhejiang University, China), Sida Peng (Zhejiang University, China), Chen Geng (Zhejiang University, China; Stanford University, USA), Linzhan Mou (Zhejiang University, China), Zihan Yan (University of Illinois Urbana-Champaign, USA), Jiaming Sun (Zhejiang University, China), Hujun Bao (Zhejiang University, China), and Xiaowei Zhou (Zhejiang University, China)

From Audio to Photoreal Embodiment: Synthesizing Humans in Conversations 1001
Evonne Ng (University of California, Berkeley; Codec Avatars Lab, Meta, Pittsburgh), Javier Romero (Codec Avatars Lab, Meta, Pittsburgh), Timur Bagautdinov (Codec Avatars Lab, Meta, Pittsburgh), Shaojie Bai (Codec Avatars Lab, Meta, Pittsburgh), Trevor Darrell (University of California, Berkeley), Angjoo Kanazawa (University of California, Berkeley), and Alexander Richard (Codec Avatars Lab, Meta, Pittsburgh)

Closely Interactive Human Reconstruction with Proxemics and Physics-Guided Adaption 1011
Buzhen Huang (National University of Singapore, Singapore; Southeast University, China), Chen Li (National University of Singapore, Singapore), Chongyang Xu (Sichuan University, China), Liang Pan (Southeast University, China), Yangang Wang (Southeast University, China), and Gim Hee Lee (National University of Singapore, Singapore)

Video-Based Human Pose Regression via Decoupled Space-Time Aggregation 1022
Jijie He (Zhejiang Gongshang University, China) and Wenwu Yang (Zhejiang Gongshang University, China)

Rethinking Generalizable Face Anti-spoofing via Hierarchical Prototype-guided Distribution Refinement in Hyperbolic Space 1032
Chengyang Hu (Shanghai Jiao Tong University, China), Ke-Yue Zhang (Tencent, China), Taiping Yao (Tencent, China), Shouhong Ding (Tencent, China), and Lizhuang Ma (Shanghai Jiao Tong University, China)

MoML: Online Meta Adaptation for 3D Human Motion Prediction 1042
Xiaoning Sun (Nanjing University of Science and Technology, China), Huaijiang Sun (Nanjing University of Science and Technology, China), Bin Li (Tianjin AiForward Science and Technology Co., Ltd., China), Dong Wei (Nanjing University of Science and Technology, China), Weiqing Li (Nanjing University of Science and Technology, China), and Jianfeng Lu (Nanjing University of Science and Technology, China)

KITRO: Refining Human Mesh by 2D Clues and Kinematic-tree Rotation 1052
Fengyuan Yang (National University of Singapore, Singapore), Kerui Gu (National University of Singapore, Singapore), and Angela Yao (National University of Singapore, Singapore)

Guess The Unseen: Dynamic 3D Scene Reconstruction from Partial 2D Glimpses 1062
Inhee Lee (Seoul National University), Byungjun Kim (Seoul National University), and Hanbyul Joo (Seoul National University)

PEGASUS: Personalized Generative 3D Avatars with Composable Attributes 1072
Hyunsoo Cha (Seoul National University, South Korea), Byungjun Kim (Seoul National University, South Korea), and Hanbyul Joo (Seoul National University, South Korea)

SDPose: Tokenized Pose Estimation via Circulation-Guide Self-Distillation 1082
Sichen Chen (Shanghai Jiao Tong University, China), Yingyi Zhang (Tencent Youtu Lab, China), Siming Huang (Tencent Youtu Lab, China), Ran Yi (Shanghai Jiao Tong University, China), Ke Fan (Shanghai Jiao Tong University, China), Ruixin Zhang (Tencent Youtu Lab, China), Peixian Chen (Tencent Youtu Lab, China), Jun Wang (Tencent Youtu Lab, China), Shouhong Ding (Tencent Youtu Lab, China), and Lizhuang Ma (Shanghai Jiao Tong University, China)

Mocap Everyone Everywhere: Lightweight Motion Capture With Smartwatches and a Head-Mounted Camera 1091
Jiye Lee (Seoul National University, South Korea) and Hanbyul Joo (Seoul National University, South Korea)

DPMesh: Exploiting Diffusion Prior for Occluded Human Mesh Recovery 1101
Yixuan Zhu (Tsinghua University, China), Ao Li (Tsinghua University, China), Yansong Tang (Tsinghua University, China), Wenliang Zhao (Tsinghua University, China), Jie Zhou (Tsinghua University, China), and Jiwen Lu (Tsinghua University, China)

DPHMs: Diffusion Parametric Head Models for Depth-based Tracking 1111
Jiapeng Tang (Technical University of Munich), Angela Dai (Technical University of Munich), Yinyu Nie (Technical University of Munich), Lev Markhasin (Sony Semiconductor Solutions Europe), Justus Thies (Technical University of Darmstadt), and Matthias Nießner (Technical University of Munich)

KTPFormer: Kinematics and Trajectory Prior Knowledge-Enhanced Transformer for 3D Human Pose Estimation 1123
Jihua Peng (The Hong Kong Polytechnic University, Hong Kong; Laboratory for Artificial Intelligence in Design, Hong Kong), Yanghong Zhou (The Hong Kong Polytechnic University, Hong Kong), and P. Y. Mok (The Hong Kong Polytechnic University, Hong Kong. Laboratory for Artificial Intelligence in Design, Hong Kong)

Exploiting Style Latent Flows for Generalizing Deepfake Video Detection 1133
Jongwook Choi (Chung-Ang Univ.), Taehoon Kim (Chung-Ang Univ, Korea), Yonghyun Jeong (Image Vision, NAVER Cloud, Korea), Seungryul Baek (AI Graduate School, UNIST, Korea), and Jongwon Choi (Chung-Ang Univ, Korea)

EMAGE: Towards Unified Holistic Co-Speech Gesture Generation via Expressive Masked Audio Gesture Modeling 1144
Haiyang Liu (The University of Tokyo), Zihao Zhu (Keio University), Giorgio Becherini (Max Planck Institute for Intelligent Systems), Yichen Peng (Japan Advanced Institute of Science and Technology), Mingyang Su (Tsinghua University), You Zhou (Independent Researcher), Xuefei Zhe (Independent Researcher), Naoya Iwamoto (Independent Researcher), Bo Zheng (Independent Researcher), and Michael J. Black (Max Planck Institute for Intelligent Systems)

A Unified Framework for Human-centric Point Cloud Video Understanding 1155
Yiteng Xu (ShanghaiTech University), Kecheng Ye (ShanghaiTech University), Xiao Han (ShanghaiTech University), Yiming Ren (ShanghaiTech University), Xinge Zhu (The Chinese University of Hong Kong), and Yuexin Ma (ShanghaiTech University)

ASH: Animatable Gaussian Splats for Efficient and Photoreal Human Rendering 1165
Haokai Pang (Max Planck Institute for Informatics, Saarland Informatics Campus, Germany; ETH Zürich, Switzerland), Heming Zhu (Max Planck Institute for Informatics, Saarland Informatics Campus, Germany), Adam Kortylewski (Max Planck Institute for Informatics, Saarland Informatics Campus, Germany; Universität Freiburg, Germany), Christian Theobalt (Max Planck Institute for Informatics, Saarland Informatics Campus, Germany; Saarbrücken Research Center for Visual Computing, Interaction and AI, Germany;), and Marc Habermann (Max Planck Institute for Informatics, Saarland Informatics Campus, Germany; Saarbrücken Research Center for Visual Computing, Interaction and AI, Germany)

CLOAF: CoLlisiOn-Aware Human Flow 1176
Andrey Davydov (EPFL, Switzerland), Martin Engilberge (EPFL, Switzerland), Mathieu Salzmann (EPFL, Switzerland), and Pascal Fua (EPFL, Switzerland)

EventEgo3D: 3D Human Motion Capture from Egocentric Event Streams 1186
Christen Millerdurai (Max Planck Institute for Informatics; Saarland University), Hiroyasu Akada (Max Planck Institute for Informatics), Jian Wang (Max Planck Institute for Informatics), Diogo Luvizon (Max Planck Institute for Informatics), Christian Theobalt (Max Planck Institute for Informatics), and Vladislav Golyanik (Max Planck Institute for Informatics)

A Call to Reflect on Evaluation Practices for Age Estimation: Comparative Analysis of the State-of-the-Art and a Unified Benchmark 1196
Jakub Paplhám (Czech Technical University in Prague, Czech Republic) and Vojtěch Franc (Czech Technical University in Prague, Czech Republic)

Holoported Characters: Real-time Free-viewpoint Rendering of Humans from Sparse RGB Cameras 1206
Ashwath Shetty (Max Planck Institute for Informatics), Marc Habermann (Max Planck Institute for Informatics), Guoxing Sun (Max Planck Institute for Informatics), Diogo Luvizon (Max Planck Institute for Informatics), Vladislav Golyanik (Max Planck Institute for Informatics), and Christian Theobalt (Max Planck Institute)

Synergistic Global-space Camera and Human Reconstruction from Videos 1216
Yizhou Zhao (Carnegie Mellon University, USA), Tuanfeng Yang Wang (Adobe Research, UK), Bhiksha Raj (Carnegie Mellon University, USA), Min Xu (Carnegie Mellon University, USA), Jimei Yang (Adobe Research, USA), and Chun-Hao Paul Huang (Adobe Research, UK)

3D Face Tracking from 2D Video through Iterative Dense UV to Image Flow 1227
Felix Taubner (LG Electronics), Prashant Raina (LG Electronics), Mathieu Tuli (LG Electronics), Eu Wern Teh (LG Electronics), Chul Lee (LG Electronics), and Jinmiao Huang (LG Electronics)

UltrAvatar: A Realistic Animatable 3D Avatar Diffusion Model with Authenticity Guided Textures 1238
Mingyuan Zhou (Innopeak Technology), Rakib Hyder (Innopeak Technology), Ziwei Xuan (Innopeak Technology), and Guojun Qi (Westlake University)

OmniMotionGPT: Animal Motion Generation with Limited Data 1249
Zhangsihao Yang (Arizona State University, USA), Mingyuan Zhou (OPPO Seattle Research Center, USA), Mengyi Shan (University of Washington, USA), Bingbing Wen (University of Washington, USA), Ziwei Xuan (OPPO Seattle Research Center, USA), Mitch Hill (OPPO Seattle Research Center, USA), Junjie Bai (OPPO Seattle Research Center, USA), Guo-Jun Qi (OPPO Seattle Research Center, USA; Westlake University, China), and Yalin Wang (Arizona State University, USA)

Text-Guided 3D Face Synthesis - From Generation to Editing 1260
Yunjie Wu (Netease Fuxi AI Lab, China), Yapeng Meng (Tsinghua University, China), Zhipeng Hu (Netease Fuxi AI Lab, China), Lincheng Li (Netease Fuxi AI Lab, China), Haoqian Wu (Netease Fuxi AI Lab, China), Kun Zhou (Zhejiang University, China), Weiwei Xu (Zhejiang University, China), and Xin Yu (University of Queensland, Australia)

Multi-scale Dynamic and Hierarchical Relationship Modeling for Facial Action Units Recognition 1270
Zihan Wang (Shenzhen University), Siyang Song (University of Leicester), Cheng Luo (Monash University), Songhe Deng (Shenzhen University), Weicheng Xie (Shenzhen University), and Linlin Shen (Shenzhen University)

LiveHPS: LiDAR-based Scene-level Human Pose and Shape Estimation in Free Environment 1281
Yiming Ren (ShanghaiTech University), Xiao Han (ShanghaiTech University), Chengfeng Zhao (ShanghaiTech University), Jingya Wang (ShanghaiTech University), Lan Xu (ShanghaiTech University), Jingyi Yu (ShanghaiTech University), and Yuexin Ma (ShanghaiTech University)

FaceChain-ImagineID: Freely Crafting High-Fidelity Diverse Talking Faces from Disentangled Audio 1292
Chao Xu (Alibaba Group), Yang Liu (Alibaba Group), Jiazheng Xing (FaceChain Community), Weida Wang (FaceChain Community), Mingze Sun (FaceChain Community), Jun Dan (FaceChain Community), Tianxin Huang (National University of Singapore), Siyuan Li (FaceChain Community), Zhi-Qi Cheng (Carnegie Mellon University), Ying Tai (Nanjing University), and Baigui Sun (Alibaba Group)

OpticalDR: A Deep Optical Imaging Model for Privacy-Protective Depression Recognition 1303
Yuchen Pan (Harbin Institute of Technology, China), Junjun Jiang (Harbin Institute of Technology, China), Kui Jiang (Harbin Institute of Technology, China), Zhihao Wu (Harbin Institute of Technology, China), Keyuan Yu (Harbin Institute of Technology, China), and Xianming Liu (Harbin Institute of Technology, China)

SCE-MAE: Selective Correspondence Enhancement with Masked Autoencoder for Self-Supervised Landmark Estimation 1313
Kejia Yin (University of Toronto, Canada), Varshanth Rao (ModiFace, Canada), Ruowei Jiang (ModiFace, Canada), Xudong Liu (University of Toronto, Canada), Parham Aarabi (University of Toronto, Canada), and David B. Lindell (University of Toronto, Canada)

TokenHMR: Advancing Human Mesh Recovery with a Tokenized Pose Representation 1323
Sai Kumar Dwivedi (Max Planck Institute for Intelligent Systems, Germany), Yu Sun (Meshcapade, Germany), Priyanka Patel (Max Planck Institute for Intelligent Systems, Germany), Yao Feng (Max Planck Institute for Intelligent Systems, Germany; Meshcapade, Germany; ETH Zurich, Switzerland), and Michael J. Black (Max Planck Institute for Intelligent Systems, Germany)

Optimizing Diffusion Noise Can Serve As Universal Motion Priors 1334
Korrawe Karunratanakul (ETH Zurich), Konpat Preechakul (University of California, Berkeley), Emre Aksan (Google), Thabo Beeler (Google), Supasorn Suwajanakorn (VISTEC), and Siyu Tang (ETH Zurich)

M&M VTO: Multi-Garment Virtual Try-On and Editing 1346
Luyang Zhu (University of Washington, USA), Yingwei Li (Google Research, USA), Nan Liu (Google Research, USA), Hao Peng (Google Research, USA), Dawei Yang (Google Research, USA), and Ira Kemelmacher-Shlizerman (Google Research, USA)

AvatarGPT: All-in-One Framework for Motion Understanding, Planning, Generation and Beyond 1357
Zixiang Zhou (Xiaobing.ai), Yu Wan (Xiaobing.ai), and Baoyuan Wang (Xiaobing.ai)

A Simple Baseline for Efficient Hand Mesh Reconstruction 1367
Zhishan Zhou (Jiiov Technology, China), Shihao Zhou (Jiiov Technology,China), Zhi Lv (Jiiov Technology, China), Minqiang Zou (Jiiov Technology, China), Yao Tang (Jiiov Technology, China), and Jiajun Liang (Jiiov Technology, China)

VINECS: Video-based Neural Character Skinning 1377
Zhouyingcheng Liao (The University of Hong Kong, Hong Kong; Max Planck Institute for Informatics, Saarland Informatics Campus, Germany), Vladislav Golyanik (Max Planck Institute for Informatics, Saarland Informatics Campus, Germany), Marc Habermann (Max Planck Institute for Informatics, Saarland Informatics Campus, Germany; Saarbruecken Research Center for Visual Computing, Interaction and AI, Germay), and Christian Theobalt (Max Planck Institute for Informatics, Saarland Informatics Campus, Germany; Saarbruecken Research Center for Visual Computing, Interaction and AI, Germay)

ConvoFusion: Multi-Modal Conversational Diffusion for Co-Speech Gesture Synthesis 1388
Muhammad Hamza Mughal (SIC, Max Planck Institute for Informatics, Germany), Rishabh Dabral (SIC, Max Planck Institute for Informatics, Germany), Ikhsanul Habibie (SIC, Max Planck Institute for Informatics, Germany), Lucia Donatelli (Vrije Universiteit Amsterdam, Netherlands), Marc Habermann (SIC, Max Planck Institute for Informatics, Germany), and Christian Theobalt (SIC, Max Planck Institute for Informatics, Germany)

Programmable Motion Generation for Open-Set Motion Control Tasks 1399
Hanchao Liu (Tsinghua University, China), Xiaohang Zhan (Tencent AI Lab, China), Shaoli Huang (Tencent AI Lab, China), Tai-Jiang Mu (Tsinghua University, China), and Ying Shan (Tencent AI Lab, China)

From Feature to Gaze: A Generalizable Replacement of Linear Layer for Gaze Estimation 1409
Yiwei Bao (Beihang University) and Feng Lu (Beihang University)

Unsupervised Gaze Representation Learning from Multi-view Face Images 1419
Yiwei Bao (Beihang University) and Feng Lu (Beihang University)

Joint2Human: High-Quality 3D Human Generation via Compact Spherical Embedding of 3D Joints.... 1429
Muxin Zhang (Tianjin University, China), Qiao Feng (Tianjin University, China), Zhuo Su (ByteDance, China), Chao Wen (ByteDance, China), Zhou Xue (Li Auto, China), and Kun Li (Tianjin University, China)

DiffHuman: Probabilistic Photorealistic 3D Reconstruction of Humans 1439
Akash Sengupta (University of Cambridge, Google Research), Thiemo Alldieck (Google Research), Nikos Kolotouros (Google Research), Enric Corona (Google Research), Andrei Zanfir (Google Research), and Cristian Sminchisescu (Google Research)

Bi-Causal: Group Activity Recognition via Bidirectional Causality 1450
Youliang Zhang (Wuhan University, China), Wenxuan Liu (Wuhan University of Technology, China), Danni Xu (National University Of Singapore, Singapore), Zhuo Zhou (Wuhan University), and Zheng Wang (Wuhan University, China)

HumanNeRF-SE: A Simple yet Effective Approach to Animate HumanNeRF with Diverse Poses . 1460
Caoyuan Ma (Wuhan University, China), Yu-Lun Liu (National Yang Ming Chiao Tung University, China), Zhixiang Wang (The University of Tokyo, Japan, National Institute of Informatics, Japan), Wu Liu (University of Science and Technology of China, China), Xinchen Liu (JD Explore Academy, China), and Zheng Wang (Wuhan University, China)

LPSNet: End-to-End Human Pose and Shape Estimation with Lensless Imaging 1471
Haoyang Ge (Tianjin University), Qiao Feng (Tianjin University), Hailong Jia (Tianjin University), Xiongzheng Li (Tianjin University), Xiangjun Yin (Tianjin University), You Zhou (Nanjing University), Jingyu Yang (Tianjin University), and Kun Li (Tianjin University)

MagicAnimate: Temporally Consistent Human Image Animation using Diffusion Model 1481
Zhongcong Xu (National University of Singapore, Singapore), Jianfeng Zhang (ByteDance, Singapore), Jun Hao Liew (ByteDance, Singapore), Hanshu Yan (ByteDance, Singapore), Jia-Wei Liu (National University of Singapore, Singapore), Chenxu Zhang (ByteDance, USA), Jiashi Feng (ByteDance, Singapore), and Mike Zheng Shou (National University of Singapore, Singapore)

RTMO: Towards High-Performance One-Stage Real-Time Multi-Person Pose Estimation 1491
Peng Lu (Tsinghua University, China), Tao Jiang (Shanghai AI Laboratory, China), Yining Li (Shanghai AI Laboratory, China), Xiangtai Li (Nanyang Technological University, Singapore), Kai Chen (Shanghai AI Laboratory, China), and Wenming Yang (Tsinghua University, China)

Human Motion Prediction Under Unexpected Perturbation 1501
Jiangbei Yue (University of Leeds, United Kingdom), Baiyi Li (University of Leeds, United Kingdom), Julien Pettré (INRIA, France), Armin Seyfried (Forschungszentrum Jülich, Germany), and He Wang (University College London, United Kingdom)

Cross-view and Cross-pose Completion for 3D Human Understanding 1512
Matthieu Armando (Naver Labs Europe), Salma Galaaoui (Naver Labs Europe), Fabien Baradel (Naver Labs Europe), Thomas Lucas (Naver Labs Europe), Vincent Leroy (Naver Labs Europe), Romain Brégier (Naver Labs Europe), Philippe Weinzaepfel (Naver Labs Europe), and Grégory Rogez (Naver Labs Europe)

Lodge: A Coarse to Fine Diffusion Network for Long Dance Generation Guided by the Characteristic Dance Primitives 1524
Ronghui Li (Tsinghua University), YuXiang Zhang (Tsinghua University), Yachao Zhang (Tsinghua University), Hongwen Zhang (Beijing Normal University), Jie Guo (Peng Cheng Laboratory), Yan Zhang (Meshcapade), Yebin Liu (Tsinghua University), and Xiu Li (Tsinghua University)

GALA: Generating Animatable Layered Assets from a Single Scan 1535
Taeksoo Kim (Seoul National University), Byungjun Kim (Seoul National University), Shunsuke Saito (Codec Avatars Lab, Meta), and Hanbyul Joo (Seoul National University)

MMM: Generative Masked Motion Model 1546
Ekkasit Pinyoanuntapong (University of North Carolina at Charlotte), Pu Wang (University of North Carolina at Charlotte), Minwoo Lee (University of North Carolina at Charlotte), and Chen Chen (University of Central Florida)

What Do You See in Vehicle? Comprehensive Vision Solution for In-Vehicle Gaze Estimation 1556
Yihua Cheng (University of Birmingham, UK), Yaning Zhu (Huazhong University of Science and Technology, China), Zongji Wang (Chinese Academy of Sciences, China), Hongquan Hao (CalmCar, China), Yongwei Liu (Calmcar, China), Shiqing Cheng (Calmcar, China), Xi Wang (Calmcar, China), and Hyung Jin Chang (University of Birmingham, UK)

Towards Variable and Coordinated Holistic Co-Speech Motion Generation 1566
Yifei Liu (South China University of Technology, China), Qiong Cao (JD Explore Academy, China), Yandong Wen (Max Planck Institute for Intelligent Systems, Germany), Huaiguang Jiang (South China University of Technology, China), and Changxing Ding (South China University of Technology, China)

Text2HOI: Text-guided 3D Motion Generation for Hand-Object Interaction 1577
Junuk Cha (Ulsan National Institute of Science and Technology, South Korea), Jihyeon Kim (Korea Electronics Technology Institute, South Korea), Jae Shin Yoon (Adobe Research, USA), and Seungryul Baek (Ulsan National Institute of Science and Technology, South Korea)

Garment Recovery with Shape and Deformation Priors 1586
Ren Li (EPFL, Switzerland), Corentin Dumery (EPFL, Switzerland), Benoît Guillard (EPFL, Switzerland), and Pascal Fua (EPFL, Switzerland)

Tri-Modal Motion Retrieval by Learning a Joint Embedding Space 1596
Kangning Yin (ShanghaiTech University, China), Shihao Zou (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences), Yuxuan Ge (ShanghaiTech University, China), and Zheng Tian (ShanghaiTech University, China)

SplattingAvatar: Realistic Real-Time Human Avatars with Mesh-Embedded Gaussian Splatting .. 1606
Zhijing Shao (The Hong Kong University of Science and Technology (Guangzhou)), Zhaolong Wang (Prometheus Vision Technology Co., Ltd.), Zhuang Li (Prometheus Vision Technology Co., Ltd.), Duotun Wang (The Hong Kong University of Science and Technology (Guangzhou)), Xiangru Lin (Prometheus Vision Technology Co., Ltd.), Yu Zhang (Prometheus Vision Technology Co., Ltd.), Mingming Fan (The Hong Kong University of Science and Technology (Guangzhou)), and Zeyu Wang (The Hong Kong University of Science and Technology (Guangzhou))

Multi-agent Long-term 3D Human Pose Forecasting via Interaction-aware Trajectory Conditioning 1617
Jaewoo Jeong (KAIST), Daehee Park (KAIST), and Kuk-Jin Yoon (KAIST)

HardMo: A Large-Scale Hardcase Dataset for Motion Capture 1629
Jiaqi Liao (Beijing University of Posts and Telecommunications, China), Chuanchen Luo (Institute of Automation, Chinese Academy of Sciences, China), Yinuo Du (Beijing University of Posts and Telecommunications, China), Yuxi Wang (Institute of Automation, Chinese Academy of Sciences, China), Xucheng Yin (University of Science and Technology Beijing, China), Man Zhang (Beijing University of Posts and Telecommunications, China), Zhaoxiang Zhang (Institute of Automation, Chinese Academy of Sciences, China), and Junran Peng (University of Science and Technology Beijing, China)

LAFS: Landmark-based Facial Self-supervised Learning for Face Recognition 1639
Zhonglin Sun (Queen Mary University of London, UK), Chen Feng (Queen Mary University of London, UK), Ioannis Patras (Queen Mary University of London, UK), and Georgios Tzimiropoulos (Queen Mary University of London, UK)

Motion Diversification Networks 1650
Hee Jae Kim (Boston University, USA) and Eshed Ohn-Bar (Boston University, USA)

NRDF: Neural Riemannian Distance Fields for Learning Articulated Pose Priors 1661
Yannan He (University of Tübingen, Germany; Tübingen AI Center, Germany), Garvita Tiwari (University of Tübingen, Germany; Tübingen AI Center, Germany; Max Planck Institute for Informatics, Saarland Informatics Campus, Germany), Tolga Birdal (Imperial College London, United Kindom), Jan Eric Lenssen (Max Planck Institute for Informatics, Saarland Informatics Campus, Germany), and Gerard Pons-Moll (University of Tübingen, Germany; Tübingen AI Center, Germany; Max Planck Institute for Informatics, Saarland Informatics Campus, Germany)

3D Face Reconstruction with the Geometric Guidance of Facial Part Segmentation 1672
Zidu Wang (Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), Xiangyu Zhu (Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), Tianshuo Zhang (Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), Baiqin Wang (Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), and Zhen Lei (Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China; Hong Kong Institute of Science & Innovation)

Unifying Top-down and Bottom-up Scanpath Prediction Using Transformers 1683
Zhibo Yang (Stony Brook University; Waymo LLC), Sounak Mondal (Stony Brook University), Seoyoung Ahn (Stony Brook University), Ruoyu Xue (Stony Brook University), Gregory Zelinsky (Stony Brook University), Minh Hoai (Stony Brook University; VinAI Research), and Dimitris Samaras (Stony Brook University)

CLIB-FIQA: Face Image Quality Assessment with Confidence Calibration 1694
Fu-Zhao Ou (City University of Hong Kong, Hong Kong SAR, China), Chongyi Li (Nankai University, Tianjin, China; NKIARI (Shenzhen-Futian), Shenzhen, China), Shiqi Wang (City University of Hong Kong, Hong Kong SAR, China), and Sam Kwong (Lingnan University, Hong Kong SAR, China)

MoST: Motion Style Transformer Between Diverse Action Contents 1705
Boeun Kim (Korea Electronics Technology Institute, Republic of Korea; Seoul National University, Republic of Korea; University of Birmingham, UK), Jungho Kim (Korea Electronics Technology Institute, Republic of Korea), Hyung Jin Chang (University of Birmingham, UK), and Jin Young Choi (Seoul National University, Republic of Korea)

TexVocab: Texture Vocabulary-conditioned Human Avatars 1715
Yuxiao Liu (Tsinghua University, China), Zhe Li (Tsinghua University, China), Yebin Liu (Tsinghua University, China), and Haoqian Wang (Tsinghua University, China)

Forecasting of 3D Whole-body Human Poses with Grasping Objects 1726
Haitao Yan (Fudan University), Qiongjie Cui (Nanjing University of Science and Technology), Jiexin Xie (Fudan University), and Shijie Guo (Fudan University)

Scaling Up Dynamic Human-Scene Interaction Modeling 1737
Nan Jiang (Peking University), Zhiyuan Zhang (Tsinghua University), Hongjie Li (Peking University), Xiaoxuan Ma (Peking University), Zan Wang (Beijing Institute of Technology), Yixin Chen (Beijing Institute for General Artificial Intelligence), Tengyu Liu (Beijing Institute for General Artificial Intelligence), Yixin Zhu (Peking University), and Siyuan Huang (Beijing Institute for General Artificial Intelligence)

Design2Cloth: 3D Cloth Generation from 2D Masks 1748
Jiali Zheng (Imperial College London), Rolandos Alexandros Potamias (Imperial College London), and Stefanos Zafeiriou (Imperial College London)

ReGenNet: Towards Human Action-Reaction Synthesis 1759
Liang Xu (Shanghai Jiao Tong University, China), Yizhou Zhou (WeChat, Tencent Inc., China), Yichao Yan (Shanghai Jiao Tong University, China), Xin Jin (Eastern Institute of Technology, Ningbo, China), Wenhan Zhu (Shanghai Jiao Tong University, China), Fengyun Rao (WeChat, Tencent Inc., China), Xiaokang Yang (Shanghai Jiao Tong University, China), and Wenjun Zeng (Eastern Institute of Technology, Ningbo, China)

MoSAR: Monocular Semi-Supervised Model for Avatar Reconstruction using Differentiable Shading 1770
Abdallah Dib (Ubisoft LaForge, Canada), Luiz Gustavo Hafemann (Ubisoft LaForge, Canada), Emeline Got (Ubisoft LaForge, Canada), Trevor Anderson (Ubisoft LaForge, Canada), Amin Fadaeinejad (York University, Canada), Rafael M. O. Cruz (Ecole de Technologie Supérieure, Canada), and Marc-André Carbonneau (Ubisoft LaForge, Canada)

FaceLift: Semi-supervised 3D Facial Landmark Localization 1781
David Ferman (Flawless AI), Pablo Garrido (Flawless AI), and Gaurav Bharaj (Flawless AI)

Fast Adaptation for Human Pose Estimation via Meta-Optimization 1792
Shengxiang Hu (Nanjing University of Science and Technology, China), Huaijiang Sun (Nanjing University of Science and Technology, China), Bin Li (Tianjin AiForward Science and Technology Co., Ltd., China), Dong Wei (Nanjing University of Science and Technology, China), Weiqing Li (Nanjing University of Science and Technology, China), and Jianfeng Lu (Nanjing University of Science and Technology, China)

FlashAvatar: High-fidelity Head Avatar with Efficient Gaussian Embedding 1802
Jun Xiang (University of Science and Technology of China, China), Xuan Gao (University of Science and Technology of China, China), Yudong Guo (University of Science and Technology of China, China), and Juyong Zhang (University of Science and Technology of China, China)

AAMDM: Accelerated Auto-regressive Motion Diffusion Model 1813
Tianyu Li (Georgia Institute of Technology, USA), Calvin Qiao (University of British Columbia, Canada), Guanqiao Ren (Beihang University, China), KangKang Yin (Simon Fraser University, Canada), and Sehoon Ha (Georgia Institute of Technology, USA)

SynSP: Synergy of Smoothness and Precision in Pose Sequences Refinement 1824
Tao Wang (Beijing University of Posts and Telecommunications, China), Lei Jin (Beijing University of Posts and Telecommunications, China), Zheng Wang (Wuhan University, China), Jianshu Li (Ant Group, China), Liang Li (Institute of computing technology Chinese Academy of Sciences, China), Fang Zhao (Nanjing University, China), Yu Cheng (National University of Singapore, Singapore), Li Yuan (Peking University, China), Li Zhou (Peking University, China), Junliang Xing (Tsinghua University, China), and Jian Zhao (China Telecom Institute of AI, China; Northwestern Polytechnical University, China)

AiOS: All-in-One-Stage Expressive Human Pose and Shape Estimation 1834
Qingping Sun (SenseTime Research, City University of Hong Kong), Yanjun Wang (SenseTime Research), Ailing Zeng (International Digital Economy Academy (IDEA)), Wanqi Yin (SenseTime Research), Chen Wei (SenseTime Research), Wenjia Wang (Shanghai AI Laboratory), Haiyi Mei (SenseTime Research), Chi-Sing Leung (City University of Hong Kong), Ziwei Liu (S-Lab, Nanyang Technological University), Lei Yang (SenseTime Research, Shanghai AI Laboratory), and Zhongang Cai (SenseTime Research, S-Lab, Nanyang Technological University, Shanghai AI Laboratory)

HumanRef: Single Image to 3D Human Generation via Reference-Guided Diffusion 1844
Jingbo Zhang (City University of Hong Kong), Xiaoyu Li (Tencent AI Lab), Qi Zhang (Tencent AI Lab), Yanpei Cao (Tencent AI Lab), Ying Shan (Tencent AI Lab), and Jing Liao (City University of Hong Kong)

Generating Human Motion in 3D Scenes from Text Descriptions 1855
Zhi Cen (Zhejiang University, China), Huaijin Pi (Zhejiang University, China), Sida Peng (Zhejiang University, China), Zehong Shen (Zhejiang University, China), Minghui Yang (Ant Group, China), Shuai Zhu (Ant Group, China), Hujun Bao (Zhejiang University, China), and Xiaowei Zhou (Zhejiang University, China)

Locally Adaptive Neural 3D Morphable Models 1867
Michail Tarasiou (Imperial College London, UK), Rolandos Alexandros Potamias (Imperial College London, UK), Eimear O'Sullivan (Imperial College London, UK), Stylianos Ploumpis (Imperial College London, UK), and Stefanos Zafeiriou (Imperial College London, UK)

IntrinsicAvatar: Physically Based Inverse Rendering of Dynamic Humans from Monocular Videos via Explicit Ray Tracing 1877
Shaofei Wang (ETH Zurich, Switzerland; University of Tuebingen, Germany; Tuebingen AI Center, Germany), Bozidar Antic (University of Tuebingen, Germany; Tuebingen AI Center, Germany), Andreas Geiger (University of Tuebingen, Germany; Tuebingen AI Center, Germany), and Siyu Tang (ETH Zurich, Switzerland)

Dynamic Inertial Poser (DynaIP): Part-Based Motion Dynamics Learning for Enhanced Human Pose Estimation with Sparse Inertial Sensors 1889
Yu Zhang (Shanghai Jiao Tong University), Songpengcheng Xia (Shanghai Jiao Tong University), Lei Chu (University of Southern California), Jiarui Yang (Shanghai Jiao Tong University), Qi Wu (Shanghai Jiao Tong University), and Ling Pei (Shanghai Jiao Tong University)

MoMask: Generative Masked Modeling of 3D Human Motions 1900
Chuan Guo (University of Alberta, Canada), Yuxuan Mu (University of Alberta, Canada), Muhammad Gohar Javed (University of Alberta, Canada), Sen Wang (University of Alberta, Canada), and Li Cheng (University of Alberta, Canada)

G-HOP: Generative Hand-Object Prior for Interaction Reconstruction and Grasp Synthesis 1911
Yufei Ye (Carnegie Mellon University, USA), Abhinav Gupta (Carnegie Mellon University, USA), Kris Kitani (Carnegie Mellon University, USA; Meta AI, USA), and Shubham Tulsiani (Carnegie Mellon University, USA)

Dynamic Support Information Mining for Category-Agnostic Pose Estimation 1921
Pengfei Ren (Beijing University of Posts and Telecommunications, China), Yuanyuan Gao (Beijing University of Posts and Telecommunications, China), Haifeng Sun (Beijing University of Posts and Telecommunications, China), Qi Qi (Beijing University of Posts and Telecommunications, China), Jingyu Wang (Beijing University of Posts and Telecommunications, China), and Jianxin Liao (Beijing University of Posts and Telecommunications, China)

Gaussian Head Avatar: Ultra High-fidelity Head Avatar via Dynamic Gaussians 1931
Yuelang Xu (Tsinghua University, China), Benwang Chen (Tsinghua University, China), Zhe Li (Tsinghua University, China), Hongwen Zhang (Beijing Normal University, China), Lizhen Wang (Tsinghua University, China), Zerong Zheng (NNKosmos Technology), and Yebin Liu (Tsinghua University, China)

Emotional Speech-driven 3D Body Animation via Disentangled Latent Diffusion 1942
Kiran Chhatre (KTH Royal Institute of Technology, Sweden), Radek Daněček (Max Planck Institute for Intelligent Systems, Germany), Nikos Athanasiou (Max Planck Institute for Intelligent Systems, Germany), Giorgio Becherini (Max Planck Institute for Intelligent Systems, Germany), Christopher Peters (KTH Royal Institute of Technology, Sweden), Michael J. Black (Max Planck Institute for Intelligent Systems, Germany), and Timo Bolkart (Max Planck Institute for Intelligent Systems, Germany)

ProxyCap: Real-time Monocular Full-body Capture in World Space via Human-Centric Proxy-to-Motion Learning 1954
Yuxiang Zhang (Tsinghua University), Hongwen Zhang (Beijing Normal University), Liangxiao Hu (Harbin Institute of Technology), Jiajun Zhang (Beijing University of Posts and Telecommunications), Hongwei Yi (Max Planck Institute for Intelligent Systems), Shengping Zhang (Harbin Institute of Technology), and Yebin Liu (Tsinghua University)

MAS: Multi-view Ancestral Sampling for 3D Motion Generation Using 2D Diffusion 1965
Roy Kapon (Tel Aviv University), Guy Tevet (Tel Aviv University), Daniel Cohen-Or (Tel Aviv University), and Amit H. Bermano (Tel Aviv University)

Efficient 3D Implicit Head Avatar with Mesh-anchored Hash Table Blendshapes 1975
Ziqian Bai (Google, Canada; Simon Fraser University, Canada), Feitong Tan (Google, USA), Sean Fanello (Google, USA), Rohit Pandey (Google, USA), Mingsong Dou (Google, USA), Shichen Liu (Google, USA), Ping Tan (The Hong Kong University of Science and Technology, Hong Kong), and Yinda Zhang (Google, USA)

Neural Sign Actors: A Diffusion Model for 3D Sign Language Production from Text 1985
Vasileios Baltatzis (Imperial College London, United Kingdom), Rolandos Alexandros Potamias (Imperial College London, United Kingdom), Evangelos Ververas (Imperial College London, United Kingdom), Guanxiong Sun (Queen's University Belfast, Northern Ireland), Jiankang Deng (Imperial College London, United Kingdom), and Stefanos Zafeiriou (Imperial College London, United Kingdom)

RAM-Avatar: Real-time Photo-Realistic Avatar from Monocular Videos with Full-body Control .. 1996
Xiang Deng (Tsinghua University), Zerong Zheng (NNKosmos), Yuxiang Zhang (Tsinghua University), Jingxiang Sun (Tsinghua University), Chao Xu (NNKosmos), Xiaodong Yang (Li Auto), Lizhen Wang (Tsinghua University), and Yebin Liu (Tsinghua University)

Sharingan: A Transformer Architecture for Multi-Person Gaze Following 2008
Samy Tafasca (Idiap Research Institute, Switzerland and École Polytechnique Fédérale de Lausanne, Switzerland), Anshul Gupta (Idiap Research Institute, Switzerland and École Polytechnique Fédérale de Lausanne, Switzerland), and Jean-Marc Odobez (Idiap Research Institute, Switzerland and École Polytechnique Fédérale de Lausanne, Switzerland)

Degrees of Freedom Matter: Inferring Dynamics from Point Trajectories 2018
Yan Zhang (Meshcapade), Sergey Prokudin (ETH Zurich), Marko Mihajlovic (ETH Zurich), Qianli Ma (Nvidia), and Siyu Tang (ETH Zurich)

Authentic Hand Avatar from a Phone Scan via Universal Hand Model 2029
Gyeongsik Moon (Meta), Weipeng Xu (Meta), Rohan Joshi (Meta), Chenglei Wu (Meta), and Takaaki Shiratori (Meta)

UniHuman: A Unified Model For Editing Human Images in the Wild 2039
Nannan Li (Boston University, USA), Qing Liu (Adobe Research, USA), Krishna Kumar Singh (Adobe Research, USA), Yilin Wang (Adobe Research, USA), Jianming Zhang (Adobe Research, USA), Bryan A. Plummer (Boston University, USA), and Zhe Lin (Adobe Research, USA)

BlockGCN: Redefine Topology Awareness for Skeleton-Based Action Recognition 2049
Yuxuan Zhou (University of Mannheim, Germany), Xudong Yan (City University of Macau, Macau), Zhi-Qi Cheng (Carnegie Mellon University, USA), Yan Yan (Chinese Academy of Sciences, China), Qi Dai (Microsoft Research, China), and Xian-Sheng Hua (Terminus Group, China)

GoMAvatar: Efficient Animatable Human Modeling from Monocular Video Using Gaussians-on-Mesh 2059
Jing Wen (University of Illinois Urbana-Champaign), Xiaoming Zhao (University of Illinois Urbana-Champaign), Zhongzheng Ren (University of Illinois Urbana-Champaign), Alexander G. Schwing (University of Illinois Urbana-Champaign), and Shenlong Wang (University of Illinois Urbana-Champaign)

WHAM: Reconstructing World-grounded Humans with Accurate 3D Motion 2070
Soyong Shin (Carnegie Mellon University), Juyong Kim (Carnegie Mellon University), Eni Halilaj (Carnegie Mellon University), and Michael J. Black (Max Planck Institute for Intelligent System)

Self-Supervised Facial Representation Learning with Facial Region Awareness 2081
Zheng Gao (Queen Mary University of London, United Kingdom) and Ioannis Patras (Queen Mary University of London, United Kingdom)

ChatPose: Chatting about 3D Human Pose 2093
Yao Feng (Max Planck Institute for Intelligent Systems, Germany; ETH Zurich, Switzerland; Meshcapade, Germany), Jing Lin (Tsinghua University, China; Meshcapade, Germany), Sai Kumar Dwivedi (Max Planck Institute for Intelligent Systems, Germany), Yu Sun (Meshcapade, Germany), Priyanka Patel (Max Planck Institute for Intelligent Systems, Germany), and Michael J. Black (Max Planck Institute for Intelligent Systems, Germany)

AUEditNet: Dual-Branch Facial Action Unit Intensity Manipulation with Implicit Disentanglement 2104
Shiwei Jin (University of California, San Diego), Zhen Wang (Qualcomm Technologies, Inc.), Lei Wang (Qualcomm Technologies, Inc.), Peng Liu (Qualcomm Technologies, Inc.), Ning Bi (Qualcomm Technologies, Inc.), and Truong Nguyen (University of California, San Diego)

Towards a Simultaneous and Granular Identity-Expression Control in Personalized Face Generation 2114
Renshuai Liu (Xiamen University, China), Bowen Ma (Netease Fuxi AI Lab, China), Wei Zhang (Netease Fuxi AI Lab, China), Zhipeng Hu (Netease Fuxi AI Lab, China), Changjie Fan (Netease Fuxi AI Lab, China), Tangjie Lv (Netease Fuxi AI Lab, China), Yu Ding (Netease Fuxi AI Lab, China), and Xuan Cheng (Xiamen University, China)

PoseIRM: Enhance 3D Human Pose Estimation on Unseen Camera Settings via Invariant Risk Minimization 2124
Yanlu Cai (Fudan University, China), Weizhong Zhang (Fudan University, China), Yuan Wu (Fudan University, China), and Cheng Jin (Fudan University, China; Haina Lab, China)

Rethinking Human Motion Prediction with Symplectic Integral 2134
*Haipeng Chen (Jilin University, Changchun, China), Kedi Lyu (Jilin University, Changchun, China), Zhenguang Liu (Zhejiang University, Hangzhou, China), Yifang Yin (Institute for Infocomm Research (I2R), A*STAR, Singapore), Xun Yang (University of Science and Technology of China, HeFei, China), and Yingda Lyu (Jilin University, Changchun, China)*

Multimodal Sense-Informed Forecasting of 3D Human Motions 2144
Zhenyu Lou (Zhejiang University, China), Qiongjie Cui (Nanjing University of Science and Technology, China), Haofan Wang (Carnegie Mellon University, America), Xu Tang (Xiaohongshu Inc., China), and Hong Zhou (Zhejiang University, China)

Semantics-aware Motion Retargeting with Vision-Language Models 2155
Haodong Zhang (Zhejiang University, China), Zhike Chen (Zhejiang University, China), Haocheng Xu (Zhejiang University, China), Lei Hao (Huawei Noah's Ark Lab, China), Xiaofei Wu (Huawei Noah's Ark Lab, China), Songcen Xu (Huawei Noah's Ark Lab, China), Zhensong Zhang (Huawei Noah's Ark Lab, China), Yue Wang (Zhejiang University, China), and Rong Xiong (Zhejiang University, China)

Makeup Prior Models for 3D Facial Makeup Estimation and Applications 2165
Xingchao Yang (CyberAgent, Japan; University of Tsukuba, Japan), Takafumi Taketomi (CyberAgent, Japan), Yuki Endo (University of Tsukuba, Japan), and Yoshihiro Kanamori (University of Tsukuba, Japan)

FaceCom: Towards High-fidelity 3D Facial Shape Completion via Optimization and Inpainting Guidance 2177
Yinglong Li (Beihang University, China), Hongyu Wu (Beihang University, China), Xiaogang Wang (Southwest University, China), Qingzhao Qin (Peking University School and Hospital of Stomatology, China), Yijiao Zhao (Peking University School and Hospital of Stomatology, China), Yong Wang (Peking University School and Hospital of Stomatology, China), and Aimin Hao (Beihang University, China)

When StyleGAN Meets Stable Diffusion: a W+ Adapter for Personalized Image Generation 2187
Xiaoming Li (S-Lab, Nanyang Technological University), Xinyu Hou (S-Lab, Nanyang Technological University), and Chen Change Loy (S-Lab, Nanyang Technological University)

MANUS: Markerless Grasp Capture using Articulated 3D Gaussians 2197
Chandradeep Pokhariya (International Institute of Information Technology Hyderabad), Ishaan Nikhil Shah (International Institute of Information Technology Hyderabad), Angela Xing (Brown University), Zekun Li (Brown University), Kefan Chen (Brown University), Avinash Sharma (International Institute of Information Technology Hyderabad), and Srinath Sridhar (Brown University)

Loose Inertial Poser: Motion Capture with IMU-attached Loose-Wear Jacket 2209
Chengxu Zuo (Xiamen University), Yiming Wang (Xiamen University), Lishuang Zhan (Xiamen University), Shihui Guo (Xiamen University), Xinyu Yi (Tsinghua University), Feng Xu (Tsinghua University), and Yipeng Qin (Cardiff University)

Anatomically Constrained Implicit Face Models 2220
Prashanth Chandran (DisneyResearch|Studios) and Gaspard Zoss (DisneyResearch|Studios)

DiffusionRegPose: Enhancing Multi-Person Pose Estimation using a Diffusion-Based End-to-End Regression Approach 2230
Dayi Tan (Tongji University), Hansheng Chen (Stanford University), Wei Tian (Tongji University), and Lu Xiong (Tongji University)

A Dual-Augmentor Framework for Domain Generalization in 3D Human Pose Estimation 2240
Qucheng Peng (University of Central Florida), Ce Zheng (University of Central Florida), and Chen Chen (University of Central Florida)

RELI11D: A Comprehensive Multimodal Human Motion Dataset and Method 2250
Ming Yan (Xiamen University, China), Yan Zhang (Xiamen University, China), Shuqiang Cai (Xiamen University, China), Shuqi Fan (Xiamen University, China), Xincheng Lin (Xiamen University, China), Yudi Dai (Xiamen University, China), Siqi Shen (Xiamen University, China), Chenglu Wen (Xiamen University, China), Lan Xu (ShanghaiTech University, China), Yuexin Ma (ShanghaiTech University, China), and Cheng Wang (Xiamen University, China)

Co-Speech Gesture Video Generation via Motion-Decoupled Diffusion Model 2263
Xu He (Tsinghua University, China), Qiaochu Huang (Tsinghua University, China), Zhensong Zhang (Huawei Noah's Ark Lab, China), Zhiwei Lin (Tsinghua University, China), Zhiyong Wu (Tsinghua University, China; The Chinese University of Hong Kong, China), Sicheng Yang (Tsinghua University, China), Minglei Li (Huawei Cloud Computing Technologies Co., Ltd, China), Zhiyi Chen (Huawei Cloud Computing Technologies Co., Ltd, China), Songcen Xu (Huawei Noah's Ark Lab, China), and Xiaofei Wu (Huawei Noah's Ark Lab, China)

HandDiff: 3D Hand Pose Estimation with Diffusion on Image-Point Cloud 2274
Wencan Cheng (Sungkyunkwan University), Hao Tang (Carnegie Mellon University), Luc Van Gool (CVL, ETH Zurich), and Jong Hwan Ko (Sungkyunkwan University)

Normalizing Flows on the Product Space of SO(3) Manifolds for Probabilistic Human Pose Modeling 2285
Olaf Dünkel (Max Planck Institute for Informatics, Germany), Tim Salzmann (Technical University of Munich, Germany), and Florian Pfaff (University of Stuttgart, Germany)

Towards Robust 3D Pose Transfer with Adversarial Learning 2295
Haoyu Chen (CVMS, University of Oulu), Hao Tang (CMU, USA), Ehsan Adeli (Stanford University, USA), and Guoying Zhao (CVMS, University of Oulu; Stanford University, USA)

PhysPT: Physics-aware Pretrained Transformer for Estimating Human Dynamics from Monocular Videos 2305
Yufei Zhang (Rensselaer Polytechnic Institute, USA), Jeffrey O. Kephart (IBM Research, USA), Zijun Cui (Rensselaer Polytechnic Institute, USA; University of Southern California, USA), and Qiang Ji (Rensselaer Polytechnic Institute, USA)

HumMUSS: Human Motion Understanding using State Space Models 2318
Arnab Mondal (McGill University), Stefano Alletto (Apple), and Denis Tome (Apple)

MultiPhys: Multi-Person Physics-aware 3D Motion Estimation 2331
Nicolas Ugrinovic (Institut de Robotica i Informatica Industrial, Spain), Boxiao Pan (Stanford University, USA), Georgios Pavlakos (UT Austin, USA), Despoina Paschalidou (Stanford University, USA), Bokui Shen (Stanford University, USA), Jordi Sanchez-Riera (Institut de Robotica i Informatica Industrial, Spain), Francesc Moreno-Noguer (Institut de Robotica i Informatica Industrial, Spain), and Leonidas Guibas (Stanford University, USA)

Physics-Aware Hand-Object Interaction Denoising 2341
Haowen Luo (Tsinghua University, China), Yunze Liu (Tsinghua University, China), and Li Yi (Tsinghua University, China)

HOIST-Former: Hand-held Objects Identification, Segmentation, and Tracking in the Wild 2351
Supreeth Narasimhaswamy (Stony Brook University), Huy Anh Nguyen (Stony Brook University), Lihan Huang (Stony Brook University), and Minh Hoai (Stony Brook University)

SCULPT: Shape-Conditioned Unpaired Learning of Pose-dependent Clothed and Textured Human Meshes 2362
Soubhik Sanyal (Max Planck Institute for Intelligent Systems, Germany), Partha Ghosh (Max Planck Institute for Intelligent Systems, Germany), Jinlong Yang (Max Planck Institute for Intelligent Systems, Germany), Michael J. Black (Max Planck Institute for Intelligent Systems, Germany), Justus Thies (Max Planck Institute for Intelligent Systems, Germany), and Timo Bolkart (Max Planck Institute for Intelligent Systems, Germany)

PFStorer: Personalized Face Restoration and Super-Resolution 2372
Tuomas Varanka (University of Oulu, Finland), Tapani Toivonen (Huawei Technologies, Finland), Soumya Tripathy (Huawei Technologies, Finland), Guoying Zhao (University of Oulu, Finland), and Erman Acar (Huawei Technologies, Finland)

MS-MANO: Enabling Hand Pose Tracking with Biomechanical Constraints 2382
Pengfei Xie (Southeast University), Wenqiang Xu (Shanghai Jiao Tong University), Tutian Tang (Shanghai Jiao Tong University), Zhenjun Yu (Shanghai Jiao Tong University), and Cewu Lu (Shanghai Jiao Tong University)

BOTH2Hands: Inferring 3D Hands from Both Text Prompts and Body Dynamics 2393
Wenqian Zhang (ShanghaiTech University, China), Molin Huang (ShanghaiTech University, China), Yuxuan Zhou (ShanghaiTech University, China), Juze Zhang (ShanghaiTech University, China), Jingyi Yu (ShanghaiTech University, China), Jingya Wang (ShanghaiTech University, China), and Lan Xu (ShanghaiTech University, China)

MeshPose: Unifying DensePose and 3D Body Mesh Reconstruction 2405
Eric-Tuan Le (University College London), Antonis Kakolyris (Snap Inc.), Petros Koutras (Snap Inc.), Himmy Tam (Snap Inc.), Efstratios Skordos (Snap Inc.), George Papandreou (Snap Inc.), Riza Alp Güler (Snap Inc.), and Iasonas Kokkinos (Snap Inc.)

CustomListener: Text-guided Responsive Interaction for User-friendly Listening Head Generation 2415
Xi Liu (Meituan, China), Ying Guo (Meituan, China), Cheng Zhen (Meituan, China), Tong Li (Meituan, China), Yingying Ao (Meituan, China), and Pengfei Yan (Meituan, China)

Generalizable Face Landmarking Guided by Conditional Face Warping 2425
Jiayi Liang (Beijing Institute of Technology, China), Haotian Liu (Beijing Institute of Technology, China), Hongteng Xu (Renmin University of China, China), and Dixin Luo (Beijing Institute of Technology, China)

Skeleton-in-Context: Unified Skeleton Sequence Modeling with In-Context Learning 2436
Xinshun Wang (Sun Yat-sen University), Zhongbin Fang (Sun Yat-sen University), Xia Li (ETH Zurich), Xiangtai Li (Nanyang Technological University), Chen Chen (University of Central Florida), and Mengyuan Liu (Peking University)

A Unified and Interpretable Emotion Representation and Expression Generation 2447
Reni Paskaleva (First Private Mathematical High School, Sofia, Bulgaria), Mykyta Holubakha (INSAIT, Sofia University, Bulgaria), Andela Ilic (ETH Zurich), Saman Motamed (INSAIT, Sofia University, Bulgaria), Luc Van Gool (ETH Zurich & INSAIT), and Danda Paudel (INSAIT, Sofia University, Bulgaria)

Artist-Friendly Relightable and Animatable Neural Heads 2457
Yingyan Xu (ETH Zurich, DisneyResearch|Studios), Prashanth Chandran (DisneyResearch|Studios), Sebastian Weiss (DisneyResearch|Studios), Markus Gross (ETH Zurich, DisneyResearch|Studios), Gaspard Zoss (DisneyResearch|Studios), and Derek Bradley (DisneyResearch|Studios)

HanDiffuser: Text-to-Image Generation With Realistic Hand Appearances 2468
Supreeth Narasimhaswamy (Stony Brook University), Uttaran Bhattacharya (Adobe), Xiang Chen (Adobe), Ishita Dasgupta (Adobe), Saayan Mitra (Adobe), and Minh Hoai (Stony Brook University)

BodyMAP - Jointly Predicting Body Mesh and 3D Applied Pressure Map for People in Bed 2480
Abhishek Tandon (Carnegie Mellon University, USA), Anujraaj Goyal (Carnegie Mellon University, USA), Henry M. Clever (NVIDIA, USA), and Zackory Erickson (Carnegie Mellon University, USA)

3D Facial Expressions through Analysis-by-Neural-Synthesis 2490
George Retsinas (Athena Research Center, Greece), Panagiotis P. Filntisis (Athena Research Center, Greece), Radek Danecek (MPI for Intelligent Systems, Germany), Victoria F. Abrevaya (MPI for Intelligent Systems, Germany), Anastasios Roussos (Foundation for Research & Technology - Hellas (FORTH), Greece), Timo Bolkart (MPI for Intelligent Systems, Germany), and Petros Maragos (National Technical University of Athens, Greece)

SelfPose3d: Self-Supervised Multi-Person Multi-View 3d Pose Estimation 2502
Vinkle Srivastav (University of Strasbourg, France), Keqi Chen (University of Strasbourg, France), and Nicolas Padoy (University of Strasbourg, France)

DiffusionPoser: Real-time Human Motion Reconstruction From Arbitrary Sparse Sensors Using Autoregressive Diffusion 2513
Tom Van Wouwe (Stanford University, USA), Seunghwan Lee (Stanford University, USA), Antoine Falisse (Stanford University, USA), Scott Delp (Stanford University, USA), and C. Karen Liu (Stanford University, USA)

Learning Diffusion Texture Priors for Image Restoration 2524
Tian Ye (The Hong Kong University of Science and Technology (Guangzhou)), Sixiang Chen (The Hong Kong University of Science and Technology (Guangzhou)), Wenhao Chai (University of Washington), Zhaohu Xing (The Hong Kong University of Science and Technology (Guangzhou)), Jing Qin (Hong Kong Polytechnic University), Ge Lin (The Hong Kong University of Science and Technology (Guangzhou), The Hong Kong University of Science and Technology), and Lei Zhu (The Hong Kong University of Science and Technology (Guangzhou), The Hong Kong University of Science and Technology)

Upscale-A-Video: Temporal-Consistent Diffusion Model for Real-World Video Super-Resolution. 2535
Shangchen Zhou (Nanyang Technological University), Peiqing Yang (Nanyang Technological University), Jianyi Wang (Nanyang Technological University), Yihang Luo (Nanyang Technological University), and Chen Change Loy (Nanyang Technological University)

Enhancing Video Super-Resolution via Implicit Resampling-based Alignment 2546
Kai Xu (National University of Singapore), Ziwei Yu (National University of Singapore), Xin Wang (Huawei International Pte Ltd, Singapore), Michael Bi Mi (Huawei International Pte Ltd, Singapore), and Angela Yao (National University of Singapore)

Boosting Neural Representations for Videos with a Conditional Decoder 2556
Xinjie Zhang (The Hong Kong University of Science and Technology), Ren Yang (SenseTime Research), Dailan He (The Chinese University of Hong Kong), Xingtong Ge (Beijing Institute of Technology), Tongda Xu (Tsinghua University), Yan Wang (Tsinghua University), Hongwei Qin (SenseTime Research), and Jun Zhang (The Hong Kong University of Science and Technology)

Restoration by Generation with Constrained Priors 2567
Zheng Ding (UC San Diego), Xuaner Zhang (Adobe), Zhuowen Tu (UC San Diego), and Zhihao Xia (Adobe)

Fantastic Animals and Where to Find Them: Segment Any Marine Animal with Dual SAM 2578
Pingping Zhang (Dalian University of Technology, China), Tianyu Yan (Dalian University of Technology, China), Yang Liu (Dalian University of Technology, China), and Huchuan Lu (Dalian University of Technology, China)

Estimating Extreme 3D Image Rotations using Cascaded Attention 2588
Shay Dekel (Bar Ilan University), Yosi Keller (Bar Ilan University), and Martin Cadik (Brno University of Technology)

Learned Scanpaths Aid Blind Panoramic Video Quality Assessment 2599
Kanglong Fan (City University of Hong Kong, Hong Kong), Wen Wen (City University of Hong Kong, Hong Kong), Mu Li (Harbin Institute of Technology (Shenzhen), China), Yifan Peng (The University of Hong Kong, Hong Kong), and Kede Ma (City University of Hong Kong, Hong Kong)

Automatic Controllable Colorization via Imagination 2609
Xiaoyan Cong (Zhejiang University, China), Yue Wu (Huawei Noah's Ark Lab, China), Qifeng Chen (The Hong Kong University of Science and Technology, China), and Chenyang Lei (CAIR, HKISI-CAS, China)

Reconstruction-free Cascaded Adaptive Compressive Sensing 2620
Chenxi Qiu (Nanjing University, Nanjing, China), Tao Yue (Nanjing University, Nanjing, China), and Xuemei Hu (Nanjing University, Nanjing, China)

A Semi-supervised Nighttime Dehazing Baseline with Spatial-Frequency Aware and Realistic Brightness Constraint 2631
Xiaofeng Cong (Southeast University), Jie Gui (Southeast University), Jing Zhang (University of Sydney), Junming Hou (Southeast University), and Hao Shen (Hefei University of Technology)

AdaBM: On-the-Fly Adaptive Bit Mapping for Image Super-Resolution 2641
Cheeun Hong (Seoul National University, Korea) and Kyoung Mu Lee (Seoul National University, Korea)

Beyond Image Super-Resolution for Image Recognition with Task-Driven Perceptual Loss 2651
Jaeha Kim (Seoul National University), Junghun Oh (Seoul National University), and Kyoung Mu Lee (Seoul National University)

Boosting Image Quality Assessment through Efficient Transformer Adaptation with Local Feature Enhancement 2662
Kangmin Xu (Wuhan University, China), Liang Liao (Nanyang Technological University, Singapore), Jing Xiao (Wuhan University, China), Chaofeng Chen (Nanyang Technological University, Singapore), Haoning Wu (Nanyang Technological University, Singapore), Qiong Yan (Sensetime Research, China), and Weisi Lin (Nanyang Technological University, Singapore)

Blur-aware Spatio-temporal Sparse Transformer for Video Deblurring 2673
Huicong Zhang (Harbin Institute of Technology, China), Haozhe Xie (Nanyang Technological University, Singapore), and Hongxun Yao (Harbin Institute of Technology, China)

XFeat: Accelerated Features for Lightweight Image Matching 2682
Guilherme Potje (Universidade Federal de Minas Gerais), Felipe Cadar (Universidade Federal de Minas Gerais), André Araujo (Google Research), Renato Martins (Universite de Bourgogne\; Universite de Lorraine, LORIA, Inria), and Erickson R. Nascimento (Universidade Federal de Minas Gerais; Microsoft)

RecDiffusion: Rectangling for Image Stitching with Diffusion Models 2692
Tianhao Zhou (University of Electronic Science and Technology of China), Haipeng Li (University of Electronic Science and Technology of China), Ziyi Wang (University of Electronic Science and Technology of China), Ao Luo (Southwest Jiaotong University, Megvii Technology), Chen-Lin Zhang (4Paradigm Inc), Jiajun Li (4Paradigm Inc), Bing Zeng (University of Electronic Science and Technology of China), and Shuaicheng Liu (University of Electronic Science and Technology of China)

Unsupervised Salient Instance Detection 2702
Xin Tian (Huawei Technologies), Ke Xu (City University of Hong Kong), and Rynson Lau (City University of Hong Kong)

FINER: Flexible Spectral-bias Tuning in Implicit NEural Representation by Variable-periodic Activation Functions 2713
Zhen Liu (Nanjing University, China), Hao Zhu (Nanjing University, China), Qi Zhang (Tencent Company, China), Jingde Fu (Nanjing University, China), Weibing Deng (Nanjing University, China), Zhan Ma (Nanjing University, China), Yanwen Guo (Nanjing University, China), and Xun Cao (Nanjing University, China)

Robust Image Denoising through Adversarial Frequency Mixup 2723
Donghun Ryou (Seoul National University, South Korea), Inju Ha (Seoul National University, South Korea), Hyewon Yoo (Seoul National University, South Korea), Dongwan Kim (Seoul National University, South Korea), and Bohyung Han (Seoul National University, South Korea)

Efficient Multi-scale Network with Learnable Discrete Wavelet Transform for Blind Motion Deblurring 2733
Xin Gao (China University of Mining Technology, China; Tsinghua University, China), Tianheng Qiu (University of Science and Technology of China, China; Chinese Academy of Sciences, China; Tsinghua University, China), Xinyu Zhang (Tsinghua University, China), Hanlin Bai (China University of Mining Technology, China), Kang Liu (China University of Mining Technology, China), Xuan Huang (Chinese Academy of Sciences, China), Hu Wei (Chinese Academy of Sciences, China), Guoying Zhang (China University of Mining Technology, China), and Huaping Liu (Tsinghua University, China)

Efficient Scene Recovery Using Luminous Flux Prior 2743
Zhongyu Li (University of Science and Technology of China, China) and Lei Zhang (University of Science and Technology of China, China)

Perception-Oriented Video Frame Interpolation via Asymmetric Blending 2753
Guangyang Wu (Shanghai Jiao Tong University), Xin Tao (Kuaishou Technology), Changlin Li (SeeKoo), Wenyi Wang (University of Electronic Science and Technology of China), Xiaohong Liu (Shanghai Jiao Tong University), and Qingqing Zheng (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences)

Modular Blind Video Quality Assessment 2763
Wen Wen (City University of Hong Kong, Hong Kong SAR, China), Mu Li (The Chinese University of Hong Kong, Shenzhen, China), Yabin Zhang (ByteDance Inc. China), Yiting Liao (ByteDance Inc. China), Junlin Li (ByteDance Inc. China), Li Zhang (ByteDance Inc. China), and Kede Ma (City University of Hong Kong, Hong Kong SAR, China)

Residual Denoising Diffusion Models 2773
Jiawei Liu (State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences), Qiang Wang (Shenyang University), Huijie Fan (State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences), Yinong Wang (The University of Hong Kong), Yandong Tang (State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences), and Liangqiong Qu (The University of Hong Kong)

JDEC: JPEG Decoding via Enhanced Continuous Cosine Coefficients 2784
Woo Kyoung Han (Korea University, Republic of Korea), Sunghoon Im (DGIST, Republic of Korea), Jaedeok Kim (NVIDIA), and Kyong Hwan Jin (Korea University, Republic of Korea)

On the Robustness of Language Guidance for Low-Level Vision Tasks: Findings from Depth Estimation 2794
Agneet Chatterjee (Arizona State University), Tejas Gokhale (University of Maryland, Baltimore County), Chitta Baral (Arizona State University), and Yezhou Yang (Arizona State University)

Blur2Blur: Blur Conversion for Unsupervised Image Deblurring on Unknown Domains 2804
Bang-Dang Pham (VinAI Research, Vietnam), Phong Tran (Mohamed bin Zayed University of Artificial Intelligence, UAE), Anh Tran (VinAI Research, Vietnam), Cuong Pham (Posts & Telecommunications Institute of Technology, Vietnam), Rang Nguyen (VinAI Research, Vietnam), and Minh Hoai (University of Adelaide, Australia)

Exploring Efficient Asymmetric Blind-Spots for Self-Supervised Denoising in Real-World Scenarios 2814
Shiyan Chen (Peking University, China), Jiyuan Zhang (Peking University, China), Zhaofei Yu (Peking University, China), and Tiejun Huang (Peking University, China)

Deep Equilibrium Diffusion Restoration with Parallel Sampling 2824
Jiezhang Cao (ETH Zurich, Switzerland), Yue Shi (Shanghai Jiao Tong University, China), Kai Zhang (Nanjing University, China), Yulun Zhang (Shanghai Jiao Tong University, China), Radu Timofte (University of Wurzburg, Germany), and Luc Van Gool (ETH Zurich, Switzerland)

PTM-VQA: Efficient Video Quality Assessment Leveraging Diverse PreTrained Models from the Wild 2835
Kun Yuan (Kuaishou Technology, China), Hongbo Liu (Tsinghua University, China), Mading Li (Kuaishou Technology, China), Muyi Sun (School of AI, BUPT, China), Ming Sun (Kuaishou Technology, China), Jiachao Gong (Kuaishou Technology, China), Jinhua Hao (Kuaishou Technology, China), Chao Zhou (Kuaishou Technology, China), and Yansong Tang (Tsinghua University, China)

Depth Information Assisted Collaborative Mutual Promotion Network for Single Image Dehazing 2846
Yafei Zhang (Kunming University of Science and Technology), Shen Zhou (Kunming University of Science and Technology), and Huafeng Li (Kunming University of Science and Technology)

Transcending the Limit of Local Window: Advanced Super-Resolution Transformer with Adaptive Token Dictionary 2856
Leheng Zhang (University of Electronic Science and Technology of China, China), Yawei Li (ETH Zurich, Switzerland), Xingyu Zhou (University of Electronic Science and Technology of China, China), Xiaorui Zhao (University of Electronic Science and Technology of China, China), and Shuhang Gu (University of Electronic Science and Technology of China, China)

Improving Image Restoration through Removing Degradations in Textual Representations 2866
Jingbo Lin (Harbin Institute of Technology, China), Zhilu Zhang (Harbin Institute of Technology, China), Yuxiang Wei (Harbin Institute of Technology, China), Dongwei Ren (Harbin Institute of Technology, China), Dongsheng Jiang (Huawei Cloud Computing Co., Ltd., China), Qi Tian (Huawei Cloud Computing Co., Ltd., China), and Wangmeng Zuo (Harbin Institute of Technology, China)

Towards Real-World HDR Video Reconstruction: A Large-Scale Benchmark Dataset and A Two-Stage Alignment Network 2879
Yong Shu (Shanghai University, China), Liquan Shen (Shanghai University, China), Xiangyu Hu (Shanghai University, China), Mengyao Li (Shanghai University, China), and Zihao Zhou (Shanghai University, China)

Spatio-Temporal Turbulence Mitigation: A Translational Perspective 2889
Xingguang Zhang (Purdue University), Nicholas Chimitt (Purdue University), Yiheng Chi (Purdue University), Zhiyuan Mao (Samsung Research America), and Stanley H. Chan (Purdue University)

Boosting Image Restoration via Priors from Pre-trained Models 2900
Xiaogang Xu (The Chinese University of Hong Kong, Hong Kong), Shu Kong (University of Macau), Tao Hu (National University of Singapore), Zhe Liu (Zhejiang Lab), and Hujun Bao (Zhejiang University)

Misalignment-Robust Frequency Distribution Loss for Image Transformation 2910
Zhangkai Ni (Tongji University, China), Juncheng Wu (Tongji University, China), Zian Wang (Tongji University, China), Wenhan Yang (Peng Cheng Laboratory, China), Hanli Wang (Tongji University, China), and Lin Ma (Meituan, China)

CoDe: An Explicit Content Decoupling Framework for Image Restoration 2920
Enxuan Gu (Dalian University of Technology, China), Hongwei Ge (Dalian University of Technology, China), and Yong Guo (South China University of Technology, China)

DSL-FIQA: Assessing Facial Image Quality via Dual-Set Degradation Learning and Landmark-Guided Transformer 2931
Wei-Ting Chen (National Taiwan University, Taiwan/Snap Inc., USA), Gurunandan Krishnan (Snap Inc.), Qiang Gao (Snap Inc., USA), Sy-Yen Kuo (National Taiwan University, Taiwan), Sizhuo Ma (Snap Inc., USA), and Jian Wang (Snap Inc., USA)

CLIPtone: Unsupervised Learning for Text-based Image Tone Adjustment 2942
Hyeongmin Lee (POSTECH), Kyoungkook Kang (POSTECH), Jungseul Ok (POSTECH), and Sunghyun Cho (POSTECH)

Adapt or Perish: Adaptive Sparse Transformer with Attentive Feature Refinement for Image Restoration 2952
Shihao Zhou (Nankai University, Tianjin, China, Nankai International Advanced Research Institute (SHENZHEN FUTIAN), Nankai University, Shenzhen, China,), Duosheng Chen (Nankai University, China), Jinshan Pan (Nanjing University of Science and Technology, China), Jinglei Shi (Nankai University, China), and Jufeng Yang (Nankai University, Tianjin, China, Nankai International Advanced Research Institute (SHENZHEN FUTIAN), Nankai University, Shenzhen, China)

CPGA: Coding Priors-Guided Aggregation Network for Compressed Video Quality Enhancement 2964
Qiang Zhu (University of Electronic Science and Technology of China), Jinhua Hao (Kuaishou Technology), Yukang Ding (Kuaishou Technology), Yu Liu (University of Electronic Science and Technology of China), Qiao Mo (University of Electronic Science and Technology of China), Ming Sun (Kuaishou Technology), Chao Zhou (Kuaishou Technology), and Shuyuan Zhu (University of Electronic Science and Technology of China)

Learning to Control Camera Exposure via Reinforcement Learning 2975
Kyunghyun Lee (LG AI Research), Ukcheol Shin (CMU), and Byeong-Uk Lee (KRAFTON)

Real-Time Exposure Correction via Collaborative Transformations and Adaptive Sampling 2984
Ziwen Li (Huazhong University of Science and Technology, China), Feng Zhang (Huazhong University of Science and Technology, China), Meng Cao (Mohamed bin Zayed University of Artificial Intelligence, The United Arab Emirates), Jinpu Zhang (Huazhong University of Science and Technology, China), Yuanjie Shao (Huazhong University of Science and Technology, China), Yuehuan Wang (Huazhong University of Science and Technology, China), and Nong Sang (Huazhong University of Science and Technology, China)

Towards Progressive Multi-Frequency Representation for Image Warping 2995
Jun Xiao (The Hong Kong Polytechnic University, Hong Kong, China), Zihang Lyu (The Hong Kong Polytechnic University, Hong Kong, China), Cong Zhang (The Hong Kong Polytechnic University, Hong Kong, China), Yakun Ju (Nanyang Technological University, Singapore), Changjian Shui (Vector Institute, Canada), and Kin-Man Lam (The Hong Kong Polytechnic University, Hong Kong, China)

HIR-Diff: Unsupervised Hyperspectral Image Restoration Via Improved Diffusion Models 3005
Li Pang (Xi'an Jiaotong University, China), Xiangyu Rui (Xi'an Jiaotong University, China), Long Cui (Xi'an Jiaotong University, China), Hongzhong Wang (Xi'an Jiaotong University, China), Deyu Meng (Xi'an Jiaotong University, China), and Xiangyong Cao (Xi'an Jiaotong University, China)

ZERO-IG: Zero-Shot Illumination-Guided Joint Denoising and Adaptive Enhancement for Low-Light Images 3015
Yiqi Shi (Harbin Engineering University, China), Duo Liu (Harbin Engineering University, China), Liguo Zhang (Harbin Engineering University,China), Ye Tian (Xidian University, China), Xuezhi Xia (Harbin Engineering University, China; Wuhan Digital Engineering Research Institute, China), and Xiaojing Fu (Harbin Engineering University, China)

Masked and Shuffled Blind Spot Denoising for Real-World Images 3025
Hamadi Chihaoui (University of Bern, Switzerland) and Paolo Favaro (University of Bern, Switzerland)

Continuous Optical Zooming: A Benchmark for Arbitrary-Scale Image Super-Resolution in Real World 3035
Huiyuan Fu (Beijing University of Posts and Telecommunications, China), Fei Peng (Beijing University of Posts and Telecommunications, China), Xianwei Li (Beijing University of Posts and Telecommunications, China), Yejun Li (Beijing University of Posts and Telecommunications, China), Xin Wang (Stony Brook University), and Huadong Ma (Beijing University of Posts and Telecommunications, China)

Laplacian-guided Entropy Model in Neural Codec with Blur-dissipated Synthesis 3045
Atefeh Khoshkhahtinat (West Virginia University, USA), Ali Zafari (West Virginia University, USA), Piyush M. Mehta (West Virginia University, USA), and Nasser M. Nasrabadi (West Virginia University, USA)

SD2Event:Self-supervised Learning of Dynamic Detectors and Contextual Descriptors for Event Cameras 3055
Yuan Gao (University of Science and Technology of China, China), Yuqing Zhu (University of Science and Technology of China, China), Xinjun Li (University of Science and Technology of China, China), Yimin Du (University of Science and Technology of China, China), and Tianzhu Zhang (University of Science and Technology of China, China)

LLaFS: When Large Language Models Meet Few-Shot Segmentation 3065
Lanyun Zhu (Singapore University of Technology and Design), Tianrun Chen (Zhejiang University), Deyi Ji (Alibaba Group), Jieping Ye (Alibaba Group), and Jun Liu (Singapore University of Technology and Design)

Telling Left from Right: Identifying Geometry-Aware Semantic Correspondence 3076
Junyi Zhang (Shanghai Jiao Tong University, China), Charles Herrmann (Google Research, USA), Junhwa Hur (Google Research, USA), Eric Chen (University of Illinois Urbana-Champaign, USA), Varun Jampani (Stability AI, USA), Deqing Sun (Google Research, USA), and Ming-Hsuan Yang (University of California at Merced, USA)

One-Shot Open Affordance Learning with Foundation Models 3086
Gen Li (University of Edinburgh), Deqing Sun (Google Research), Laura Sevilla-Lara (University of Edinburgh), and Varun Jampani (Stability AI)

CorrMatch: Label Propagation via Correlation Matching for Semi-Supervised Semantic Segmentation 3097
Boyuan Sun (Nankai University, China), Yuqi Yang (Nankai University, China), Le Zhang (University of Electronic Science and Technology, China), Ming-Ming Cheng (Nankai University, China), and Qibin Hou (Nankai University, China)

Collaborating Foundation Models for Domain Generalized Semantic Segmentation 3108
Yasser Benigmim (Institut Polytechnique de Paris, France), Subhankar Roy (University of Aberdeen, United Kingdom), Slim Essid (Institut Polytechnique de Paris, France), Vicky Kalogeiton (Institut Polytechnique de Paris, France), and Stéphane Lathuilière (Institut Polytechnique de Paris, France)

FocSAM: Delving Deeply into Focused Objects in Segmenting Anything 3120
You Huang (Xiamen University, China), Zongyu Lan (Xiamen University, China), Liujuan Cao (Xiamen University, China), Xianming Lin (Xiamen University, China), Shengchuan Zhang (Xiamen University, China), Guannan Jiang (Contemporary Amperex Technology Co. Limited, China), and Rongrong Ji (Xiamen University, China)

Finsler-Laplace-Beltrami Operators with Application to Shape Analysis 3131
Simon Weber (Technical University of Munich), Thomas Dagès (Technion - Israel Institute of Technology), Maolin Gao (Technical University of Munich), and Daniel Cremers (Technical University of Munich)

Neural Implicit Representation for Building Digital Twins of Unknown Articulated Objects 3141
Yijia Weng (Stanford University, USA), Bowen Wen (NVIDIA, USA), Jonathan Tremblay (NVIDIA, USA), Valts Blukis (NVIDIA, USA), Dieter Fox (NVIDIA, USA), Leonidas Guibas (Stanford University, USA), and Stan Birchfield (NVIDIA, USA)

Putting the Object Back into Video Object Segmentation 3151
Ho Kei Cheng (University of Illinois Urbana-Champaign), Seoung Wug Oh (Adobe Research), Brian Price (Adobe Research), Joon-Young Lee (Adobe Research), and Alexander Schwing (University of Illinois Urbana-Champaign)

BA-SAM: Scalable Bias-Mode Attention Mask for Segment Anything Model 3162
Yiran Song (Shanghai Jiao Tong University, China), Qianyu Zhou (Shanghai Jiao Tong University, China), Xiangtai Li (Nanyang Technological University, Singapore), Deng-Ping Fan (Nankai International Advanced Research Institute (SHENZHEN FUTIAN), China; Nankai University, China), Xuequan Lu (La Trobe University, Australia), and Lizhuang Ma (Shanghai Jiao Tong University, China)

Task-aligned Part-aware Panoptic Segmentation through Joint Object-Part Representations 3174
Daan de Geus (Eindhoven University of Technology, The Netherlands) and Gijs Dubbelman (Eindhoven University of Technology, The Netherlands)

Open-World Semantic Segmentation Including Class Similarity 3184
Matteo Sodano (University of Bonn), Federico Magistri (University of Bonn), Lucas Nunes (University of Bonn), Jens Behley (University of Bonn), and Cyrill Stachniss (University of Bonn)

Hierarchical Histogram Threshold Segmentation – Auto-terminating High-detail Oversegmentation 3195
Thomas V. Chang (Nuremberg Institute of Technology, Germany), Simon Seibt (Nuremberg Institute of Technology, Germany), and Bartosz von Rymon Lipinski (Nuremberg Institute of Technology, Germany)

AlignSAM: Aligning Segment Anything Model to Open Context via Reinforcement Learning 3205
Duojun Huang (Sun Yat-sen University, China), Xinyu Xiong (Sun Yat-sen University, China), Jie Ma (Sun Yat-sen University, China), Jichang Li (Sun Yat-sen University, China), Zequn Jie (Meituan, China), Lin Ma (Meituan, China), and Guanbin Li (Sun Yat-sen University, China)

SANeRF-HQ: Segment Anything for NeRF in High Quality 3216
Yichen Liu (The Hong Kong University of Science and Technology, Hong Kong SAR), Benran Hu (Carnegie Mellon University, USA), Chi-Keung Tang (The Hong Kong University of Science and Technology, Hong Kong SAR), and Yu-Wing Tai (Dartmouth College, USA)

UniVS: Unified and Universal Video Segmentation with Prompts as Queries 3227
Minghan Li (The Hong Kong Polytechnic University), Shuai Li (The Hong Kong Polytechnic University), Xindong Zhang (OPPO Research Institute), and Lei Zhang (The Hong Kong Polytechnic University)

RankED: Addressing Imbalance and Uncertainty in Edge Detection Using Ranking-based Losses 3239
Bedrettin Cetinkaya (Middle East Technical University, Turkey), Sinan Kalkan (Middle East Technical University, Turkey), and Emre Akbas (Middle East Technical University, Turkey)

Event-assisted Low-Light Video Object Segmentation 3250
Hebei Li (University of Science and Technology of China), Jin Wang (University of Science and Technology of China), Jiahui Yuan (University of Science and Technology of China), Yue Li (University of Science and Technology of China), Wenming Weng (University of Science and Technology of China), Yansong Peng (University of Science and Technology of China), Yueyi Zhang (University of Science and Technology of China), Zhiwei Xiong (University of Science and Technology of China), and Xiaoyan Sun (University of Science and Technology of China; Institute of Artificial Intelligence, Hefei Comprehensive National Science Center)

Density-Guided Semi-Supervised 3D Semantic Segmentation with Dual-Space Hardness Sampling 3260
Jianan Li (University of Chinese Academy of Sciences, China) and Qiulei Dong (University of Chinese Academy of Sciences, China)

Exploring Regional Clues in CLIP for Zero-Shot Semantic Segmentation 3270
Yi Zhang (Beihang University), Meng-Hao Guo (Tsinghua University), Miao Wang (Beihang University), and Shi-Min Hu (Tsinghua University)

Category-Level Multi-Part Multi-Joint 3D Shape Assembly 3281
Yichen Li (MIT CSAIL), Kaichun Mo (NVIDIA), Yueqi Duan (Tsinghua University), He Wang (Peking University), Jiequan Zhang (Stanford University), and Lin Shao (National University of Singapore)

SAI3D: Segment Any Instance in 3D Scenes 3292
Yingda Yin (Peking University, China), Yuzheng Liu (Peking University, China), Yang Xiao (Ecole des Ponts ParisTech, France), Daniel Cohen-Or (Tel-Aviv University, Israel), Jingwei Huang (Tencent, China), and Baoquan Chen (Peking University, China)

Towards the Uncharted: Density-Descending Feature Perturbation for Semi-supervised Semantic Segmentation 3303
Xiaoyang Wang (Xi'an Jiaotong Liverpool University, China), Huihui Bai (Beijing Jiaotong University, China), Limin Yu (Xi'an Jiaotong Liverpool University, China), Yao Zhao (Beijing Jiaotong University, China), and Jimin Xiao (Xi'an Jiaotong Liverpool University, China)

Hybrid Functional Maps for Crease-Aware Non-Isometric Shape Matching 3313
Lennart Bastian (Technical University of Munich, Germany), Yizheng Xie (Technical University of Munich, Germany), Nassir Navab (Technical University of Munich, Germany), and Zorah Lähner (University of Siegen, Germany; University of Bonn, Germany; Lamarr Institute, Germany)

Hunting Attributes: Context Prototype-Aware Learning for Weakly Supervised Semantic Segmentation 3324
Feilong Tang (Monash University), Zhongxing Xu (Cornell University), Zhaojun Qu (Xi'an Jiaotong-Liverpool University), Wei Feng (Monash University), Xingjian Jiang (University of Michigan), and Zongyuan Ge (Monash University)

Self-Calibrating Vicinal Risk Minimisation for Model Calibration 3335
Jiawei Liu (The Australian National University), Changkun Ye (The Australian National University), Ruikai Cui (The Australian National University), and Nick Barnes (The Australian National University)

ECLIPSE: Efficient Continual Learning in Panoptic Segmentation with Visual Prompt Tuning 3346
Beomyoung Kim (NAVER Cloud; KAIST), Joonsang Yu (NAVER Cloud), and Sung Ju Hwang (KAIST)

Clustering Propagation for Universal Medical Image Segmentation 3357
Yuhang Ding (University of Technology Sydney), Liulei Li (University of Technology Sydney), Wenguan Wang (Zhejiang University), and Yi Yang (Zhejiang University)

Addressing Background Context Bias in Few-Shot Segmentation through Iterative Modulation ... 3370
Lanyun Zhu (Singapore University of Technology and Design, Singapore), Tianrun Chen (Zhejiang University, China), Jianxiong Yin (NVIDIA AI Tech Centre, Singapore), Simon See (NVIDIA AI Tech Centre, Singapore), and Jun Liu (Singapore University of Technology and Design, Singapore)

Cross-Domain Few-Shot Segmentation via Iterative Support-Query Correspondence Mining 3380
Jiahao Nie (Nanyang Technological University, Singapore), Yun Xing (Nanyang Technological University, Singapore), Gongjie Zhang (Black Sesame Technologies, Singapore), Pei Yan (Huazhong University of Science and Technology, China; Nanyang Technological University, Singapore), Aoran Xiao (Nanyang Technological University, Singapore), Yap-Peng Tan (Nanyang Technological University, Singapore), Alex C. Kot (Nanyang Technological University, Singapore), and Shijian Lu (Nanyang Technological University, Singapore)

RankMatch: Exploring the Better Consistency Regularization for Semi-supervised Semantic Segmentation 3391
Huayu Mai (University of Science and Technology of China), Rui Sun (University of Science and Technology of China), Tianzhu Zhang (University of Science and Technology of China; Deep Space Exploration Lab), and Feng Wu (University of Science and Technology of China)

QDFormer: Towards Robust Audiovisual Segmentation in Complex Environments with Quantization-based Semantic Decomposition 3402
Xiang Li (Carnegie Mellon University), Jinglu Wang (Microsoft Research Asia), Xiaohao Xu (University of Michigan, Ann Arbor), Xiulian Peng (Microsoft Research Asia), Rita Singh (Carnegie Mellon University), Yan Lu (Microsoft Research Asia), and Bhiksha Raj (Carnegie Mellon University)

Frequency-Adaptive Dilated Convolution for Semantic Segmentation 3414
Linwei Chen (Beijing Institute of Technology), Lin Gu (The University of Tokyo), Dezhi Zheng (Beijing Institute of Technology), and Ying Fu (Beijing Institute of Technology)

SED: A Simple Encoder-Decoder for Open-Vocabulary Semantic Segmentation 3426
Bin Xie (Tianjin University, China), Jiale Cao (Tianjin University, China), Jin Xie (Chongqing University, China), Fahad Shahbaz Khan (Mohamed bin Zayed University of Artificial Intelligence, UAE; Linköping University, Sweden), and Yanwei Pang (Tianjin University, China; Shanghai Artificial Intelligence Laboratory, China)

PSDPM: Prototype-based Secondary Discriminative Pixels Mining for Weakly Supervised Semantic Segmentation 3437
Xinqiao Zhao (XJTLU, China), Ziqian Yang (XJTLU, China), Tianhong Dai (University of Aberdeen, UK), Bingfeng Zhang (China University of Petroleum, China), and Jimin Xiao (XJTLU, China)

Coupled Laplacian Eigenmaps for Locally-Aware 3D Rigid Point Cloud Matching 3447
Matteo Bastico (Mines Paris, Université PSL, France), Etienne Decencière (Mines Paris, Université PSL, France), Laurent Corté (Mines Paris, Université PSL, France), Yannick Tillier (Mines Paris, Université PSL, France), and David Ryckelynck (Mines Paris, Université PSL, France)

Universal Segmentation at Arbitrary Granularity with Language Instruction 3459
Yong Liu (Tsinghua Shenzhen International Graduate School, Tsinghua University), Cairong Zhang (ByteDance Inc.), Yitong Wang (ByteDance Inc.), Jiahao Wang (The University of Hong Kong), Yujiu Yang (Tsinghua Shenzhen International Graduate School, Tsinghua University), and Yansong Tang (Tsinghua Shenzhen International Graduate School, Tsinghua University)

PartDistill: 3D Shape Part Segmentation by Vision-Language Model Distillation 3470
Ardian Umam (National Yang Ming Chiao Tung University, Taiwan), Cheng-Kun Yang (MediaTek), Min-Hung Chen (NVIDIA), Jen-Hui Chuang (National Yang Ming Chiao Tung University, Taiwan), and Yen-Yu Lin (National Yang Ming Chiao Tung University, Taiwan)

HIT: Estimating Internal Human Implicit Tissues from the Body Surface 3480
Marilyn Keller (Max Planck Institute for Intelligent Systems, Tuebingen, Germany), Vaibhav Arora (Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, France), Abdelmouttaleb Dakri (Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, France), Shivam Chandhok (Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, France), Jürgen Machann (IDM, Helmholtz Center Munich at the University of T ubingen), Andreas Fritsche (IDM, Helmholtz Center Munich at the University of T ubingen), Michael J. Black (Max Planck Institute for Intelligent Systems, Tuebingen, Germany), and Sergi Pujades (Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, France)

Open-Vocabulary Segmentation with Semantic-Assisted Calibration 3491
Yong Liu (Shenzhen Key Laboratory of Ubiquitous Data Enabling, Shenzhen International Graduate School), Sule Bai (Shenzhen Key Laboratory of Ubiquitous Data Enabling, Shenzhen International Graduate School), Guanbin Li (Sun Yat-sen University), Yitong Wang (ByteDance Inc.), and Yansong Tang (Shenzhen Key Laboratory of Ubiquitous Data Enabling, Shenzhen International Graduate School)

GraCo: Granularity-Controllable Interactive Segmentation 3501
Yian Zhao (Peking University, China), Kehan Li (Peking University, China), Zesen Cheng (Peking University, China), Pengchong Qiao (Peking University, China), Xiawu Zheng (Xiamen University, China), Rongrong Ji (Xiamen University, China), Chang Liu (Tsinghua University, China), Li Yuan (Peking University, China), and Jie Chen (Peking University, China)

Unleashing the Potential of SAM for Medical Adaptation via Hierarchical Decoding 3511
Zhiheng Cheng (East China Normal University, China), Qingyue Wei (Stanford University, USA), Hongru Zhu (Johns Hopkins University, USA), Yan Wang (East China Normal University, China), Liangqiong Qu (The University of Hong Kong, Hong Kong), Wei Shao (University of Florida, USA), and Yuyin Zhou (UC Santa Cruz, USA)

EAGLE: Eigen Aggregation Learning for Object-Centric Unsupervised Semantic Segmentation ... 3523
Chanyoung Kim (Yonsei University, Republic of Korea), Woojung Han (Yonsei University, Republic of Korea), Dayun Ju (Yonsei University, Republic of Korea), and Seong Jae Hwang (Yonsei University, Republic of Korea)

DuPL: Dual Student with Trustworthy Progressive Learning for Robust Weakly Supervised Semantic Segmentation 3534
Yuanchen Wu (Shanghai University, China), Xichen Ye (Shanghai University, China), Kequan Yang (Shanghai University, China), Jide Li (Shanghai University, China), and Xiaoqiang Li (Shanghai University, China)

Vanishing-Point-Guided Video Semantic Segmentation of Driving Scenes 3544
Diandian Guo (ETH Zürich, Switzerland), Deng-Ping Fan (Nankai University, China), Tongyu Lu (ETH Zürich, Switzerland), Christos Sakaridis (ETH Zürich, Switzerland), and Luc Van Gool (ETH Zürich, Switzerland)

Diffuse, Attend, and Segment: Unsupervised Zero-Shot Segmentation using Stable Diffusion 3554
Junjiao Tian (Georgia Institute of Technology, USA), Lavisha Aggarwal (Google), Andrea Colaco (Google), Zsolt Kira (Georgia Institute of Technology, USA), and Mar Gonzalez-Franco (Google)

ODIN: A Single Model for 2D and 3D Segmentation 3564
Ayush Jain (Carnegie Mellon University, USA), Pushkal Katara (Carnegie Mellon University, USA), Nikolaos Gkanatsios (Carnegie Mellon University, USA), Adam W. Harley (Stanford University, USA), Gabriel Sarch (Carnegie Mellon University, USA), Kriti Aggarwal (Microsoft, USA), Vishrav Chaudhary (Microsoft, USA), and Katerina Fragkiadaki (Carnegie Mellon University, USA)

Infer from What You Have Seen Before: Temporally-dependent Classifier for Semi-supervised Video Segmentation 3575
Jiafan Zhuang (Shantou University), Zilei Wang (University of Science and Technology of China), Yixin Zhang (University of Science and Technology of China), and Zhun Fan (Shantou University)

Semantic-aware SAM for Point-Prompted Instance Segmentation 3585
Zhaoyang Wei (University of Chinese Academy of Sciences, China), Pengfei Chen (University of Chinese Academy of Sciences, China), Xuehui Yu (University of Chinese Academy of Sciences, China), Guorong Li (University of Chinese Academy of Sciences, China), Jianbin Jiao (University of Chinese Academy of Sciences, China), and Zhenjun Han (University of Chinese Academy of Sciences, China)

Class Tokens Infusion for Weakly Supervised Semantic Segmentation 3595
Sung-Hoon Yoon (KAIST), Hoyong Kwon (KAIST), Hyeonseong Kim (KAIST), and Kuk-Jin Yoon (KAIST)

Separate and Conquer: Decoupling Co-occurrence via Decomposition and Representation for Weakly Supervised Semantic Segmentation 3606
Zhiwei Yang (Fudan University), Kexue Fu (Shandong Computer Science Center), Minghong Duan (Fudan University), Linhao Qu (Fudan University), Shuo Wang (Fudan University), and Zhijian Song (Fudan University)

Style Blind Domain Generalized Semantic Segmentation via Covariance Alignment and Semantic Consistence Contrastive Learning 3616
Woo-Jin Ahn (Korea University, Republic of Korea), Geun-Yeong Yang (Korea University, Republic of Korea), Hyun-Duck Choi (Chonnam National University), and Myo-Taeg Lim (Korea University, Republic of Korea)

AllSpark: Reborn Labeled Features from Unlabeled in Transformer for Semi-Supervised Semantic Segmentation 3627
Haonan Wang (The Hong Kong University of Science and Technology, Hong Kong SAR), Qixiang Zhang (The Hong Kong University of Science and Technology, Hong Kong SAR), Yi Li (The Hong Kong University of Science and Technology, Hong Kong SAR), and Xiaomeng Li (The Hong Kong University of Science and Technology, Hong Kong SAR)

Unsupervised Semantic Segmentation Through Depth-Guided Feature Correlation and Sampling 3637
Leon Sick (Ulm University, Germany), Dominik Engel (Ulm University, Germany), Pedro Hermosilla (TU Vienna, Austria), and Timo Ropinski (Ulm University, Germany)

PoNQ: a Neural QEM-based Mesh Representation 3647
Nissim Maruani (Inria, Université Côte d'Azur), Maks Ovsjanikov (LIX, Ecole polytechnique, IP Paris), Pierre Alliez (Inria, Université Côte d'Azur), and Mathieu Desbrun (Inria Saclay, Ecole polytechnique)

Spectral Meets Spatial: Harmonising 3D Shape Matching and Interpolation 3658
Dongliang Cao (University of Bonn), Marvin Eisenberger (Technical University of Munich), Nafie El Amrani (University of Bonn), Daniel Cremers (Technical University of Munich), and Florian Bernard (University of Bonn)

CosalPure: Learning Concept from Group Images for Robust Co-Saliency Detection 3669
Jiayi Zhu (East China Normal University, China), Qing Guo (IHPC &CFAR,Agency for Science, Technology and Research, Singapore), Felix Juefei-Xu (New York University, USA), Yihao Huang (Nanyang Technological University, Singapore), Yang Liu (Nanyang Technological University, Singapore), and Geguang Pu (East China Normal University, China; Shanghai Industrial Control Safety Innovation Tech. Co., Ltd, China)

ContextSeg: Sketch Semantic Segmentation by Querying the Context with Attention 3679
Jiawei Wang (Shandong University, China) and Changjian Li (The University of Edinburgh, United Kingdom)

Training-Free Open-Vocabulary Segmentation with Offline Diffusion-Augmented Prototype Generation 3689
Luca Barsellotti (University of Modena and Reggio Emilia, Italy), Roberto Amoroso (University of Modena and Reggio Emilia, Italy), Marcella Cornia (University of Modena and Reggio Emilia, Italy), Lorenzo Baraldi (University of Modena and Reggio Emilia, Italy), and Rita Cucchiara (University of Modena and Reggio Emilia, IIT-CNR, Italy)

ASAM: Boosting Segment Anything Model with Adversarial Tuning 3699
Bo Li (vivo Mobile Communication Co., Ltd, China), Haoke Xiao (vivo Mobile Communication Co., Ltd, China), and Lv Tang (vivo Mobile Communication Co., Ltd, China)

In-Context Matting 3711
He Guo (Huazhong University of Science and Technology), Zixuan Ye (Huazhong University and Science and Technology), Zhiguo Cao (Huazhong University and Science and Technology), and Hao Lu (Huazhong University and Science and Technology)

Weakly Supervised Point Cloud Semantic Segmentation via Artificial Oracle 3721
Hyeokjun Kweon (Korea Advanced Institute of Science and Technology), Jihun Kim (Korea Advanced Institute of Science and Technology), and Kuk-Jin Yoon (Korea Advanced Institute of Science and Technology)

Contextrast: Contextual Contrastive Learning for Semantic Segmentation 3732
Changki Sung (Korea Advanced Institute of Science and Technology, Korea), Wanhee Kim (Kookmin University, Korea), Jungho An (Kookmin University, Korea), Wooju Lee (Korea Advanced Institute of Science and Technology, Korea), Hyungtae Lim (Korea Advanced Institute of Science and Technology, Korea), and Hyun Myung (Korea Advanced Institute of Science and Technology, Korea)

Parameter Efficient Fine-tuning via Cross Block Orchestration for Segment Anything Model 3743
Zelin Peng (Shanghai Jiao Tong University), Zhengqin Xu (Shanghai Jiao Tong University), Zhilin Zeng (Shanghai Jiao Tong University), Lingxi Xie (Huawei Inc.), Qi Tian (Huawei Inc.), and Wei Shen (Shanghai Jiao Tong University)

CADTalk: An Algorithm and Benchmark for Semantic Commenting of CAD Programs 3753
Haocheng Yuan (The University of Edinburgh, United Kindom), Jing Xu (The University of Edinburgh, United Kindom), Hao Pan (Microsoft Research Asia, China), Adrien Bousseau (Inria, Universite Cote d'Azur; Delft University of Technology; France), Niloy J. Mitra (University College London; Adobe Research; United Kindom), and Changjian Li (The University of Edinburgh)

Point2CAD: Reverse Engineering CAD Models from 3D Point Clouds 3763
Yujia Liu (ETH Zurich, Switzerland), Anton Obukhov (ETH Zurich, Switzerland), Jan Dirk Wegner (University of Zurich, Switzerland), and Konrad Schindler (ETH Zurich, Switzerland)

Rethinking Interactive Image Segmentation with Low Latency, High Quality, and Diverse Prompts 3773
Qin Liu (The University of North Carolina at Chapel Hill), Jaemin Cho (The University of North Carolina at Chapel Hill), Mohit Bansal (The University of North Carolina at Chapel Hill), and Marc Niethammer (The University of North Carolina at Chapel Hill)

General Object Foundation Model for Images and Videos at Scale 3783
Junfeng Wu (Huazhong University of Science and Technology, China), Yi Jiang (ByteDance, China), Qihao Liu (Johns Hopkins University, USA), Zehuan Yuan (ByteDance, China), Xiang Bai (Huazhong University of Science and Technology, China), and Song Bai (ByteDance, Singapore)

Frozen CLIP: A Strong Backbone for Weakly Supervised Semantic Segmentation 3796
Bingfeng Zhang (China University of Petroleum (East China)), Siyue Yu (Xi'an Jiaotong-Liverpool University), Yunchao Wei (Beijing Jiaotong University), Yao Zhao (Beijing Jiaotong University), and Jimin Xiao (Xi'an Jiaotong-Liverpool University)

Guided Slot Attention for Unsupervised Video Object Segmentation 3807
Minhyeok Lee (Yonsei University), Suhwan Cho (Yonsei University), Dogyoon Lee (Yonsei University), Chaewon Park (Yonsei University), Jungho Lee (Yonsei University), and Sangyoun Lee (Yonsei University)

Unlocking the Potential of Pre-trained Vision Transformers for Few-Shot Semantic Segmentation through Relationship Descriptors 3817
Ziqin Zhou (Australian Institute for Machine Learning (AIML), University of Adelaide, Australia), Hai-Ming Xu (Australian Institute for Machine Learning (AIML), University of Adelaide, Australia), Yangyang Shu (Australian Institute for Machine Learning (AIML), University of Adelaide, Australia), and Lingqiao Liu (Australian Institute for Machine Learning (AIML), University of Adelaide, Australia)

Grounding Everything: Emerging Localization Properties in Vision-Language Transformers 3828
Walid Bousselham (University of Bonn, Germany; Goethe University Frankfurt, Germany), Felix Petersen (Stanford University, USA), Vittorio Ferrari (Synthesia.io, England), and Hilde Kuehne (University of Bonn, Germany; Goethe University Frankfurt, Germany; MIT-IBM Watson AI Lab, USA)

No Time to Train: Empowering Non-Parametric Networks for Few-shot 3D Scene Segmentation . 3838
Xiangyang Zhu (City University of Hong Kong, Hong Kong, China), Renrui Zhang (Shanghai AI Lab, China), Bowei He (City University of Hong Kong, Hong Kong, China), Ziyu Guo (Shanghai AI Lab, China), Jiaming Liu (Peking University, China), Han Xiao (Shanghai AI Lab, China), Chaoyou Fu (Tencent Youtu Lab, China), Hao Dong (Peking University, China), and Peng Gao (Shanghai AI Lab, China)

Continual Segmentation with Disentangled Objectness Learning and Class Recognition 3848
Yizheng Gong (Xi'an Jiaotong-Liverpool University, China), Siyue Yu (Xi'an Jiaotong-Liverpool University, China), Xiaoyang Wang (Xi'an Jiaotong-Liverpool University, China), and Jimin Xiao (Xi'an Jiaotong-Liverpool University, China)

GSVA: Generalized Segmentation via Multimodal Large Language Models 3858
Zhuofan Xia (Tsinghua University, China), Dongchen Han (Tsinghua University, China), Yizeng Han (Tsinghua University, China), Xuran Pan (Tsinghua University, China), Shiji Song (Tsinghua University, China), and Gao Huang (Tsinghua University, China)

MaGGIe: Masked Guided Gradual Human Instance Matting 3870
Chuong Huynh (University of Maryland, College Park), Seoung Wug Oh (Adobe Research), Abhinav Shrivastava (University of Maryland, College Park), and Joon-Young Lee (Adobe Research)

EFormer: Enhanced Transformer towards Semantic-Contour Features of Foreground for Portraits Matting 3880
Zitao Wang (Xidian University, China), Qiguang Miao (Xidian University, China), Yue Xi (Xidian University, China), and Peipei Zhao (Xidian University, China)

Segment Any Event Streams via Weighted Adaptation of Pivotal Tokens 3890
Zhiwen Chen (Xidian University, China), Zhiyu Zhu (City University of Hong Kong), Yifan Zhang (City University of Hong Kong), Junhui Hou (City University of Hong Kong), Guangming Shi (Xidian University; Pazhou Lab), and Jinjian Wu (Xidian University; Pazhou Lab)

PolarMatte: Fully Computational Ground-Truth-Quality Alpha Matte Extraction for Images and Video using Polarized Screen Matting 3901
Kenji Enomoto (Adobe, Inc.), TJ Rhodes (Adobe, Inc.), Brian Price (Adobe, Inc.), and Gavin Miller (Adobe, Inc.)

Segment Every Out-of-Distribution Object 3910
Wenjie Zhao (University of Texas at Dallas), Jia Li (University of Texas at Dallas), Xin Dong (Harvard University), Yu Xiang (University of Texas at Dallas), and Yunhui Guo (University of Texas at Dallas)

Multi-view Aggregation Network for Dichotomous Image Segmentation 3921
Qian Yu (Dalian University of Technology, China), Xiaoqi Zhao (Dalian University of Technology, China), Youwei Pang (Dalian University of Technology, China), Lihe Zhang (Dalian University of Technology, China), and Huchuan Lu (Dalian University of Technology, China)

pix2gestalt: Amodal Segmentation by Synthesizing Wholes 3931
Ege Ozguroglu (Columbia University, USA), Ruoshi Liu (Columbia University, USA), Dídac Surís (Columbia University, USA), Dian Chen (Toyota Research Institute, USA), Achal Dave (Toyota Research Institute, USA), Pavel Tokmakov (Toyota Research Institute, USA), and Carl Vondrick (Columbia University, USA)

Rethinking Prior Information Generation with CLIP for Few-Shot Segmentation 3941
Jin Wang (China University of PetroleumEast China), Bingfeng Zhang (China University of PetroleumEast China), Jian Pang (China University of PetroleumEast China), Honglong Chen (China University of PetroleumEast China), and Weifeng Liu (China University of PetroleumEast China)

Image-to-Image Matching via Foundation Models: A New Perspective for Open-Vocabulary Semantic Segmentation 3952
Yuan Wang (University of Science and Technology of China), Rui Sun (University of Science and Technology of China), Naisong Luo (University of Science and Technology of China), Yuwen Pan (University of Science and Technology of China), and Tianzhu Zhang (University of Science and Technology of China/Deep Space Exploration Laboratory)

Domain Separation Graph Neural Networks for Saliency Object Ranking 3964
Zijian Wu (Nanjing University of Science and Technology, China), Jun Lu (Nanjing University of Science and Technology, China), Jing Han (Nanjing University of Science and Technology, China), Lianfa Bai (Nanjing University of Science and Technology, China), Yi Zhang (Nanjing University of Science and Technology, China), Zhuang Zhao (Nanjing University of Science and Technology, China), and Siyang Song (University of Leicester, UK)

DIOD: Self-Distillation Meets Object Discovery 3975
Sandra Kara (Université Paris-Saclay, France), Hejer Ammar (Université Paris-Saclay, France), Julien Denize (Université Paris-Saclay, France), Florian Chabot (Université Paris-Saclay, France), and Quoc-Cuong Pham (Université Paris-Saclay, France)

DiverGen: Improving Instance Segmentation by Learning Wider Data Distribution with More Diverse Generative Data 3986
Chengxiang Fan (Zhejiang University, China), Muzhi Zhu (Zhejiang University, China), Hao Chen (Zhejiang University, China), Yang Liu (Zhejiang University, China), Weijia Wu (Zhejiang University, China), Huaqi Zhang (vivo Mobile Communication Co.), and Chunhua Shen (Zhejiang University, China)

Rethinking Few-shot 3D Point Cloud Semantic Segmentation 3996
*Zhaochong An (University of Copenhagen, Denmark), Guolei Sun (ETH Zurich, Switzerland), Yun Liu (A*STAR, Singapore), Fayao Liu (A*STAR, Singapore), Zongwei Wu (University of Wurzburg, Germany), Dan Wang (University of Copenhagen, Denmark), Luc Van Gool (ETH Zurich, Switzerland), and Serge Belongie (University of Copenhagen, Denmark)*

Training Vision Transformers for Semi-Supervised Semantic Segmentation 4007
Xinting Hu (Max Planck Institute for Informatics, Germany), Li Jiang (Max Planck Institute for Informatics, Germany), and Bernt Schiele (Max Planck Institute for Informatics, Germany)

Open3DIS: Open-Vocabulary 3D Instance Segmentation with 2D Mask Guidance 4018
Phuc Nguyen (VinAI Research, Vietnam), Tuan Duc Ngo (UMass Amherst, USA), Evangelos Kalogerakis (UMass Amherst, USA), Chuang Gan (MIT-IBM Watson AI Lab, USA), Anh Tran (VinAI Research, Vietnam), Cuong Pham (Posts & Telecommunications Inst. of Tech., Vietnam), and Khoi Nguyen (VinAI Research, Vietnam)

Emergent Open-Vocabulary Semantic Segmentation from Off-the-shelf Vision-Language Models 4029
Jiayun Luo (Nanyang Technological University, Singapore), Siddhesh Khandelwal (University of British Columbia, Vector Institute for AI, Canada), Leonid Sigal (University of British Columbia, Vector Institute for AI, Canada), and Boyang Li (Nanyang Technological University, Singapore)

Memory-Scalable and Simplified Functional Map Learning 4041
Robin Magnet (Ecole Polytechnique, France) and Maks Ovsjanikov (Ecole Polytechnique, France)

MFP: Making Full Use of Probability Maps for Interactive Image Segmentation 4051
Chaewon Lee (Korea University), Seon-Ho Lee (Korea University), and Chang-Su Kim (Korea University)

Spherical Mask: Coarse-to-Fine 3D Point Cloud Instance Segmentation with Spherical Representation 4060
Sangyun Shin (University of Oxford, United Kingdom), Kaichen Zhou (University of Oxford, United Kingdom), Madhu Vankadari (University of Oxford, United Kingdom), Andrew Markham (University of Oxford, United Kingdom), and Niki Trigoni (University of Oxford, United Kingdom)

Adaptive Bidirectional Displacement for Semi-Supervised Medical Image Segmentation 4070
Hanyang Chi (China University of Petroleum (East China)), Jian Pang (China University of Petroleum (East China)), Bingfeng Zhang (China University of Petroleum (East China)), and Weifeng Liu (China University of Petroleum (East China))

RobustSAM: Segment Anything Robustly on Degraded Images 4081
Wei-Ting Chen (National Taiwan University, Taiwan/ Snap Inc., USA), Yu-Jiet Vong (National Taiwan University, Taiwan), Sy-Yen Kuo (National Taiwan University, Taiwan), Sizhuo Ma (Snap Inc., USA), and Jian Wang (Snap Inc., USA)

LAKE-RED: Camouflaged Images Generation by Latent Background Knowledge Retrieval-Augmented Diffusion 4092
Pancheng Zhao (Nankai University, Tianjin, China,Nankai International Advanced Research Institute (SHENZHEN FUTIAN), Nankai University, Shenzhen, China), Peng Xu (Tsinghua University, Beijing, China), Pengda Qin (Alibaba Group, China), Deng-Ping Fan (Nankai International Advanced Research Institute (SHENZHEN FUTIAN), Nankai University, Shenzhen, China,Nankai University, Tianjin, China), Zhicheng Zhang (Nankai University, Tianjin, China,Nankai International Advanced Research Institute (SHENZHEN FUTIAN), Nankai University, Shenzhen, China), Guoli Jia (Nankai University, Tianjin, China), Bowen Zhou (Tsinghua University, Beijing, China), and Jufeng Yang (Nankai University, Tianjin, China,Nankai International Advanced Research Institute (SHENZHEN FUTIAN), Nankai University, Shenzhen, China)

Learn to Rectify the Bias of CLIP for Unsupervised Semantic Segmentation 4102
Jingyun Wang (Beihang University, China) and Guoliang Kang (Beihang University, China; Zhongguancun Laboratory, China)

CAT-Seg: Cost Aggregation for Open-Vocabulary Semantic Segmentation 4113
Seokju Cho (Korea University), Heeseong Shin (Korea University), Sunghwan Hong (Korea University), Anurag Arnab (Google Research,), Paul Hongsuck Seo (Korea University), and Seungryong Kim (Korea University)

Prompt-Driven Referring Image Segmentation with Instance Contrasting 4124
Chao Shang (University of Electronic Science and Technology of China, China), Zichen Song (University of Electronic Science and Technology of China, China), Heqian Qiu (University of Electronic Science and Technology of China, China), Lanxiao Wang (University of Electronic Science and Technology of China, China), Fanman Meng (University of Electronic Science and Technology of China, China), and Hongliang Li (University of Electronic Science and Technology of China, China)

Kandinsky Conformal Prediction: Efficient Calibration of Image Segmentation Algorithms 4135
Joren Brunekreef (Netherlands Cancer Institute), Eric Marcus (Netherlands Cancer Institute), Ray Sheombarsing (Netherlands Cancer Institute), Jan-Jakob Sonke (Netherlands Cancer Institute), and Jonas Teuwen (Netherlands Cancer Institute)

OVFoodSeg: Elevating Open-Vocabulary Food Image Segmentation via Image-Informed Textual Representation 4144
Xiongwei Wu (Singapore Management University), Sicheng Yu (Singapore Management University), Ee-Peng Lim (Singapore Management University), and Chong-Wah Ngo (Singapore Management University)

Back to 3D: Few-Shot 3D Keypoint Detection with Back-Projected 2D Features 4154
Thomas Wimmer (École Polytechnique, France. Technical University of Munich, Germany), Peter Wonka (KAUST, Saudi Arabia), and Maks Ovsjanikov (École Polytechnique, France)

Deciphering 'What' and 'Where' Visual Pathways from Spectral Clustering of Layer-Distributed Neural Representations 4165
Xiao Zhang (University of Chicago, USA), David Yunis (Toyota Technological Institute at Chicago, USA), and Michael Maire (University of Chicago, USA)

Open Vocabulary Semantic Scene Sketch Understanding 4176
Ahmed Bourouis (Surrey Institute for People-Centered AI, CVSSP, University of Surrey, UK), Judith E. Fan (Stanford University, USA), and Yulia Gryaditskaya (Surrey Institute for People-Centered AI, CVSSP, University of Surrey, UK)

USE: Universal Segment Embeddings for Open-Vocabulary Image Segmentation 4187
Xiaoqi Wang (The Ohio State University, USA), Wenbin He (Bosch Research North America & Bosch Center for Artificial Intelligence (BCAI), USA), Xiwei Xuan (University of California Davis, USA), Clint Sebastian (Bosch Center for Artificial Intelligence (BCAI), Germany), Jorge Piazentin Ono (Bosch Research North America & Bosch Center for Artificial Intelligence (BCAI), USA), Xin Li (Bosch Research North America & Bosch Center for Artificial Intelligence (BCAI), USA), Sima Behpour (Bosch Research North America & Bosch Center for Artificial Intelligence (BCAI), USA), Thang Doan (Bosch Research North America & Bosch Center for Artificial Intelligence (BCAI), USA), Liang Gou (Bosch Research North America & Bosch Center for Artificial Intelligence (BCAI), USA), Han-Wei Shen (The Ohio State University, USA), and Liu Ren (Bosch Research North America & Bosch Center for Artificial Intelligence (BCAI), USA)

Diff-Plugin: Revitalizing Details for Diffusion-based Low-level Tasks 4197
Yuhao Liu (City University of Hong Kong, China), Zhanghan Ke (City University of Hong Kong, China), Fang Liu (City University of Hong Kong, China), Nanxuan Zhao (Adobe Research, USA), and Rynson W.H. Lau (City University of Hong Kong, China)

XCube: Large-Scale 3D Generative Modeling using Sparse Voxel Hierarchies 4209
Xuanchi Ren (NVIDIA, University of Toronto, Vector Institute), Jiahui Huang (NVIDIA), Xiaohui Zeng (NVIDIA, University of Toronto, Vector Institute), Ken Museth (NVIDIA), Sanja Fidler (NVIDIA, University of Toronto, Vector Institute), and Francis Williams (NVIDIA)

SC-GS: Sparse-Controlled Gaussian Splatting for Editable Dynamic Scenes 4220
Yi-Hua Huang (The University of Hong Kong), Yang-Tian Sun (The University of Hong Kong), Ziyi Yang (Zhejiang University), Xiaoyang Lyu (The University of Hong Kong), Yan-Pei Cao (VAST), and Xiaojuan Qi (The University of Hong Kong)

StyLitGAN: Image-Based Relighting via Latent Control 4231
Anand Bhattad (University of Illinois Urbana Champaign), James Soole (University of Illinois Urbana Champaign), and D.A. Forsyth (University of Illinois Urbana Champaign)

Image Sculpting: Precise Object Editing with 3D Geometry Control 4241
Jiraphon Yenphraphai (New York University, USA), Xichen Pan (New York University, USA), Sainan Liu (Intel Labs, USA), Daniele Panozzo (New York University, USA), and Saining Xie (New York University, USA)

Paint3D: Paint Anything 3D with Lighting-Less Texture Diffusion Models 4252
Xianfang Zeng (Tencent PCG, China), Xin Chen (Tencent PCG, China), Zhongqi Qi (Tencent PCG, China), Wen Liu (Tencent PCG, China), Zibo Zhao (ShanghaiTech University, China), Zhibin Wang (Tencent PCG, China), Bin Fu (Tencent PCG, China), Yong Liu (Zhejiang University, China), and Gang Yu (Tencent PCG, China)

Holo-Relighting: Controllable Volumetric Portrait Relighting from a Single Image 4263
Yiqun Mei (Johns Hopkins University), Yu Zeng (Johns Hopkins University), He Zhang (Adobe Inc.), Zhixin Shu (Adobe Inc.), Xuaner Zhang (Adobe Inc.), Sai Bi (Adobe Inc.), Jianming Zhang (Adobe Inc.), HyunJoon Jung (Adobe Inc.), and Vishal M. Patel (Johns Hopkins University)

Neural Fields as Distributions: Signal Processing Beyond Euclidean Space 4274
Daniel Rebain (University of British Columbia, Canada), Soroosh Yazdani (Google Research, USA), Kwang Moo Yi (University of British Columbia, Canada), and Andrea Tagliasacchi (Google DeepMind, USA; Simon Fraser University, Canada; University of Toronto, Canada)

TexOct: Generating Textures of 3D Models with Octree-based Diffusion 4284
Jialun Liu (Baidu Inc., China), Chenming Wu (Baidu Inc., China), Xinqi Liu (Baidu Inc., China), Xing Liu (Baidu Inc., China), Jinbo Wu (Baidu Inc., China), Haotian Peng (Baidu Inc., China), Chen Zhao (Baidu Inc., China), Haocheng Feng (Baidu Inc., China), Jingtuo Liu (Baidu Inc., China), and Errui Ding (Baidu Inc., China)

Differentiable Micro-Mesh Construction 4294
Yishun Dou (Huawei, China), Zhong Zheng (Huawei, China), Qiaoqiao Jin (Shanghai Jiao Tong University, China), Rui Shi (Shanghai Jiao Tong University, China), Yuhan Li (Shanghai Jiao Tong University, China), and Bingbing Ni (Shanghai Jiao Tong University, China)

TextureDreamer: Image-Guided Texture Synthesis Through Geometry-Aware Diffusion 4304
Yu-Ying Yeh (University of California, San Diego, USA), Jia-Bin Huang (University of Maryland, College Park, USA), Changil Kim (Meta Reality Lab, USA), Lei Xiao (Meta Reality Lab, USA), Thu Nguyen-Phuoc (Meta Reality Lab, USA), Numair Khan (Meta Reality Lab, USA), Cheng Zhang (Meta Reality Lab, USA), Manmohan Chandraker (University of California, San Diego, USA), Carl S Marshall (Meta Reality Lab, USA), Zhao Dong (Meta Reality Lab, USA), and Zhengqin Li (Meta Reality Lab, USA)

As-Plausible-As-Possible: Plausibility-Aware Mesh Deformation Using 2D Diffusion Priors 4315
Seungwoo Yoo (KAIST), Kunho Kim (KAIST), Vladimir G. Kim (Adobe Research), and Minhyuk Sung (KAIST)

Breathing Life Into Sketches Using Text-to-Video Priors 4325
Rinon Gal (Tel Aviv University, NVIDIA), Yael Vinker (Tel Aviv University), Yuval Alaluf (Tel Aviv University), Amit Bermano (Tel Aviv University), Daniel Cohen-Or (Tel Aviv University), Ariel Shamir (Reichman University), and Gal Chechik (NVIDIA)

Real-Time Neural BRDF with Spherically Distributed Primitives 4337
Yishun Dou (Huawei, China), Zhong Zheng (Huawei, China), Qiaoqiao Jin (Shanghai Jiao Tong University, China), Bingbing Ni (Shanghai Jiao Tong University, China), Yugang Chen (Huawei, China), and Junxiang Ke (Huawei, China)

Paint-it: Text-to-Texture Synthesis via Deep Convolutional Texture Map Optimization and Physically-Based Rendering 4347
Kim Youwang (Pohang University of Science and Technology, Republic of Korea), Tae-Hyun Oh (Pohang University of Science and Technology, Republic of Korea), and Gerard Pons-Moll (University of Tübingen, Germany)

Neural Super-Resolution for Real-time Rendering with Radiance Demodulation 4357
Jia Li (Shandong University, China), Ziling Chen (Shandong University, China), Xiaolong Wu (Shandong University, China), Lu Wang (Shandong University, China), Beibei Wang (Nanjing University, China), and Lei Zhang (The Hong Kong Polytechnic University, China)

DiffAvatar: Simulation-Ready Garment Optimization with Differentiable Simulation 4368
Yifei Li (Massachusetts Institute of Technology), Hsiao-yu Chen (Meta Reality Labs), Egor Larionov (Meta Reality Labs), Nikolaos Sarafianos (Meta Reality Labs), Wojciech Matusik (Massachusetts Institute of Technology), and Tuur Stuyck (Meta Reality Labs)

Material Palette: Extraction of Materials from a Single Image 4379
Ivan Lopes (Inria, France), Fabio Pizzati (University of Oxford, United Kingdom), and Raoul de Charette (Inria, France)

PhysGaussian: Physics-Integrated 3D Gaussians for Generative Dynamics 4389
Tianyi Xie (University of California, Los Angeles, USA), Zeshun Zong (University of California, Los Angeles, USA), Yuxing Qiu (University of California, Los Angeles, USA), Xuan Li (University of California, Los Angeles, USA), Yutao Feng (Zhejiang University, China), Yin Yang (University of Utah, USA), and Chenfanfu Jiang (University of California, Los Angeles, USA)

Differentiable Point-based Inverse Rendering 4399
Hoon-Gyu Chung (POSTECH, South Korea), Seokjun Choi (POSTECH, South Korea), and Seung-Hwan Baek (POSTECH, South Korea)

Towards a Perceptual Evaluation Framework for Lighting Estimation 4410
Justine Giroux (Université Laval, Canada), Mohammad Reza Karimi Dastjerdi (Université Laval, Canada), Yannick Hold-Geoffroy (Adobe Research, USA), Javier Vazquez-Corral (Universitat Autònoma de Barcelona, Spain), and Jean-François Lalonde (Université Laval, Canada)

Vector Graphics Generation via Mutually Impulsed Dual-domain Diffusion 4420
Zhongyin Zhao (Shanghai Jiao Tong University, China), Ye Chen (Shanghai Jiao Tong University, China), Zhangli Hu (Shanghai Jiao Tong University, China), Xuanhong Chen (Shanghai Jiao Tong University, China & USC-SJTU Institute of Cultural and Creative Industry), and Bingbing Ni (Shanghai Jiao Tong University, China & USC-SJTU Institute of Cultural and Creative Industry)

MatFuse: Controllable Material Generation with Diffusion Models 4429
Giuseppe Vecchio (University of Catania, Italy), Renato Sortino (University of Catania, Italy), Simone Palazzo (University of Catania, Italy), and Concetto Spampinato (University of Catania, Italy)

TexTile: A Differentiable Metric for Texture Tileability 4439
Carlos Rodriguez-Pardo (Universidad Rey Juan Carlos), Dan Casas (Universidad Rey Juan Carlos), Elena Garces (Universidad Rey Juan Carlos; SEDDI), and Jorge Lopez-Moreno (Universidad Rey Juan Carlos; SEDDI)

PIE-NeRF: Physics-based Interactive Elastodynamics with NeRF 4450
Yutao Feng (Zhejiang University, China; University of Utah, USA), Yintong Shang (University of Utah, USA), Xuan Li (University of California, Los Angeles, USA), Tianjia Shao (Zhejiang University, China), Chenfanfu Jiang (University of California, Los Angeles, USA), and Yin Yang (University of Utah, USA)

HashPoint: Accelerated Point Searching and Sampling for Neural Rendering 4462
Jiahao Ma (Australian National University, Australia), Miaomiao Liu (Australian National University, Australia), David Ahmedt-Aristizabal (CSIRO Data61, Australia), and Chuong Nguyen (CSIRO Data61, Australia)

3D Paintbrush: Local Stylization of 3D Shapes with Cascaded Score Distillation 4473
Dale Decatur (University of Chicago), Itai Lang (University of Chicago), Kfir Aberman (Snap Research), and Rana Hanocka (University of Chicago)

DUDF: Differentiable Unsigned Distance Fields with Hyperbolic Scaling 4484
Miguel Fainstein (Universidad Torcuato Di Tella, Argentina), Viviana Siless (Universidad Torcuato Di Tella, Argentina), and Emmanuel Iarussi (Universidad Torcuato Di Tella, Argentina; CONICET, Argentina)

Diffusion 3D Features (Diff3F): Decorating Untextured Shapes with Distilled Semantic Features 4494
Niladri Shekhar Dutt (University College London; Ready Player Me), Sanjeev Muralikrishnan (University College London), and Niloy J. Mitra (University College London; Adobe Research)

LeGO: Leveraging a Surface Deformation Network for Animatable Stylized Face Generation with One Example 4505
Soyeon Yoon (Korea Advanced Institute of Science and Technology, Korea), Kwan Yun (Korea Advanced Institute of Science and Technology, Korea), Kwanggyoon Seo (Korea Advanced Institute of Science and Technology, Korea), Sihun Cha (Korea Advanced Institute of Science and Technology, Korea), Jung Eun Yoo (Korea Advanced Institute of Science and Technology, Korea), and Junyong Noh (Korea Advanced Institute of Science and Technology, Korea)

Dr. Bokeh: DiffeRentiable Occlusion-aware Bokeh Rendering 4515
Yichen Sheng (Purdue University), Zixun Yu (Purdue University), Lu Ling (Purdue University), Zhiwen Cao (Adobe), Xuaner Zhang (Adobe), Xin Lu (Adobe), Ke Xian (Huazhong University of Science and Technology), Haiting Lin (Adobe), and Bedrich Benes (Purdue University)

DiffInDScene: Diffusion-based High-Quality 3D Indoor Scene Generation 4526
Xiaoliang Ju (The Chinese University of Hong Kong), Zhaoyang Huang (The Chinese University of Hong Kong), Yijin Li (Zhejiang University), Guofeng Zhang (Zhejiang University), Yu Qiao (Shanghai AI Laboratory), and Hongsheng Li (The Chinese University of Hong Kong)

LightOctree: Lightweight 3D Spatially-Coherent Indoor Lighting Estimation 4536
Xuecan Wang (Beihang University, China), Shibang Xiao (Beihang University, China), and Xiaohui Liang (Beihang University, China; Zhongguancun Laboratory, China)

SVGDreamer: Text Guided SVG Generation with Diffusion Model 4546
Ximing Xing (Beihang University), Haitao Zhou (Beihang University), Chuang Wang (Beihang University), Jing Zhang (Beihang University), Dong Xu (The University of Hong Kong), and Qian Yu (Beihang University)

Control4D: Efficient 4D Portrait Editing with Text 4556
Ruizhi Shao (Tsinghua University), Jingxiang Sun (Tsinghua University), Cheng Peng (Tsinghua University), Zerong Zheng (Tsinghua University), Boyao Zhou (Tsinghua University), Hongwen Zhang (Beijing Normal University), and Yebin Liu (Tsinghua University)

HumanNorm: Learning Normal Diffusion Model for High-quality and Realistic 3D Human Generation 4568
Xin Huang (Northwestern Polytechnical University, China), Ruizhi Shao (Tsinghua University, China), Qi Zhang (Northwestern Polytechnical University, China), Hongwen Zhang (Tsinghua University, China), Ying Feng (Northwestern Polytechnical University, China), Yebin Liu (Tsinghua University, China), and Qing Wang (Northwestern Polytechnical University, China)

Video2Game: Real-time, Interactive, Realistic and Browser-Compatible Environment from a Single Video 4578
Hongchi Xia (Shanghai Jiao Tong University, China), Zhi-Hao Lin (University of Illinois Urbana-Champaign, USA), Wei-Chiu Ma (Cornell University, USA), and Shenlong Wang (University of Illinois Urbana-Champaign, USA)

NIVeL: Neural Implicit Vector Layers for Text-to-Vector Generation 4589
Vikas Thamizharasan (University of Massachusetts Amherst, USA), Difan Liu (Adobe Research, USA), Matthew Fisher (Adobe Research, USA), Nanxuan Zhao (Adobe Research, USA), Evangelos Kalogerakis (University of Massachusetts Amherst, USA), and Michal Lukac (Adobe Research, USA)

ESR-NeRF: Emissive Source Reconstruction Using LDR Multi-view Images 4598
Jinseo Jeong (Seoul National University, Korea), Junseo Koo (Seoul National University, Korea), Qimeng Zhang (Korea University, Korea), and Gunhee Kim (Seoul National University, Korea)

DreamPropeller: Supercharge Text-to-3D Generation with Parallel Sampling 4610
Linqi Zhou (Stanford University), Andy Shih (Stanford University), Chenlin Meng (Pika Labs), and Stefano Ermon (Stanford University)

GenesisTex: Adapting Image Denoising Diffusion to Texture Space 4620
Chenjian Gao (Beihang University, China), Boyan Jiang (Tencent, China), Xinghui Li (Tencent, China), Yingpeng Zhang (Tencent, China), and Qian Yu (Beihang University, China)

Mosaic-SDF for 3D Generative Models 4630
Lior Yariv (GenAI Meta, Israel and Weizmann Institute of Science, Israel), Omri Puny (Weizmann Institute of Science, Israel), Oran Gafni (GenAI Meta, Israel), and Yaron Lipman (FAIR Meta, Israel and Weizmann Institute of Science, Israel)

NeRF Analogies: Example-Based Visual Attribute Transfer for NeRFs 4640
Michael Fischer (University College London, UK), Zhengqin Li (Meta Reality Labs, USA), Thu Nguyen-Phuoc (Meta Reality Labs, USA), Aljaz Bozic (Meta Reality Labs, USA), Zhao Dong (Meta Reality Labs, USA), Carl Marshall (Meta Reality Labs, USA), and Tobias Ritschel (University College London, UK)

Hyper-MD: Mesh Denoising with Customized Parameters Aware of Noise Intensity and Geometric Characteristics 4651
Xingtao Wang (Harbin Institute of Technology), Hongliang Wei (Harbin Institute of Technology), Xiaopeng Fan (Harbin Institute of Technology), and Debin Zhao (Harbin Institute of Technology)

QUADify: Extracting Meshes with Pixel-level Details and Materials from Images 4661
Maximilian Frühauf (ETH Zürich & DisneyResearch|Studios, Switzerland), Hayko Riemenschneider (DisneyResearch|Studios, Switzerland), Markus Gross (ETH Zürich & DisneyResearch|Studios, Switzerland), and Christopher Schroers (DisneyResearch|Studios, Switzerland)

SfmCAD: Unsupervised CAD Reconstruction by Learning Sketch-based Feature Modeling Operations 4671
Pu Li (Institute of Automation, Chinese Academy of Sciences, China), Jianwei Guo (Institute of Automation, Chinese Academy of Sciences, China), Huibin Li (Institute of Automation, Chinese Academy of Sciences, China), Bedrich Benes (Purdue University, USA), and Dong-Ming Yan (Institute of Automation, Chinese Academy of Sciences, China)

Self-Supervised Dual Contouring 4681
Ramana Sundararaman (LIX, Ecole Polytechnique), Roman Klokov (LIX, Ecole Polytechnique), and Maks Ovsjanikov (LIX, Ecole Polytechnique)

SVDTree: Semantic Voxel Diffusion for Single Image Tree Reconstruction 4692
Yuan Li (Institute of Automation, Chinese Academy of Sciences, China), Zhihao Liu (The University of Tokyo, Japan), Bedrich Benes (Purdue University, USA), Xiaopeng Zhang (Institute of Automation, Chinese Academy of Sciences, China), and Jianwei Guo (Institute of Automation, Chinese Academy of Sciences, China)

Text-Conditioned Generative Model of 3D Strand-based Human Hairstyles 4703
Vanessa Sklyarova (Max Planck Institute for Intelligent Systems, Tübingen, Germany; ETH Zürich, Switzerland), Egor Zakharov (ETH Zürich, Switzerland), Otmar Hilliges (ETH Zürich, Switzerland), Michael J. Black (Max Planck Institute for Intelligent Systems, Tübingen, Germany), and Justus Thies (Max Planck Institute for Intelligent Systems, Tübingen, Germany; Technical University of Darmstadt, Germany)

CAD-SIGNet: CAD Language Inference from Point Clouds using Layer-wise Sketch Instance Guided Attention 4713
Mohammad Sadil Khan (SnT, University of Luxembourg), Elona Dupont (SnT, University of Luxembourg), Sk Aziz Ali (German Research Center for Artificial Intelligence), Kseniya Cherenkova (Artec3D - SnT, University of Luxembourg), Anis Kacem (SnT, University of Luxembourg), and Djamila Aouada (SnT, University of Luxembourg)

Functional Diffusion 4723
Biao Zhang (KAUST, Saudi Arabia) and Peter Wonka (KAUST, Saudi Arabia)

Orals 2A Image & Video Synthesis

FreeU: Free Lunch in Diffusion U-Net 4733
Chenyang Si (Nanyang Technological University, Singapore), Ziqi Huang (Nanyang Technological University, Singapore), Yuming Jiang (Nanyang Technological University, Singapore), and Ziwei Liu (Nanyang Technological University, Singapore)

Ranni: Taming Text-to-Image Diffusion for Accurate Instruction Following 4744
Yutong Feng (Alibaba Group), Biao Gong (Alibaba Group), Di Chen (Alibaba Group), Yujun Shen (Ant Group), Yu Liu (Alibaba Group), and Jingren Zhou (Alibaba Group)

Instruct-Imagen: Image Generation with Multi-modal Instruction 4754
Hexiang Hu (Google), Kelvin C.K. Chan (Google), Yu-Chuan Su (Google), Wenhu Chen (Google), Yandong Li (Google), Kihyuk Sohn (Google), Yang Zhao (Google), Xue Ben (Google), Boqing Gong (Google), William Cohen (Google), Ming-Wei Chang (Google), and Xuhui Jia (Google)

Attention Calibration for Disentangled Text-to-Image Personalization 4764
Yanbing Zhang (ECUST, China), Mengping Yang (ECUST, China), Qin Zhou (ECUST, China), and Zhe Wang (ECUST, China)

Style Aligned Image Generation via Shared Attention 4775
Amir Hertz (Google Research, Israel), Andrey Voynov (Google Research, Israel), Shlomi Fruchter (Google Research, Israel), and Daniel Cohen-Or (Tel Aviv University, Israel)

Orals 2B Deep Learning Architectures and Techniques

Neural Redshift: Random Networks are not Random Functions 4786
Damien Teney (Idiap Research Institute, Switzerland), Armand Mihai Nicolicioiu (ETH Zurich, Switzerland), Valentin Hartmann (EPFL, Switzerland), and Ehsan Abbasnejad (University of Adelaide, Australia)

Neural Lineage 4797
Runpeng Yu (National University of Singapore) and Xinchao Wang (National University of Singapore)

Learning Structure-from-Motion with Graph Attention Networks 4808
Lucas Brynte (Chalmers University of Technology, Sweden), José Pedro Iglesias (Chalmers University of Technology, Sweden), Carl Olsson (Chalmers University of Technology, Sweden), and Fredrik Kahl (Chalmers University of Technology, Sweden)

Florence-2: Advancing a Unified Representation for a Variety of Vision Tasks 4818
Bin Xiao (Microsoft), Haiping Wu (Microsoft), Weijian Xu (Microsoft), Xiyang Dai (Microsoft), Houdong Hu (Microsoft), Yumao Lu (Microsoft), Michael Zeng (Microsoft), Ce Liu (Microsoft), and Lu Yuan (Microsoft)

In Search of a Data Transformation That Accelerates Neural Field Training 4830
Junwon Seo (POSTECH, Republic of Korea), Sangyoon Lee (POSTECH, Republic of Korea), Kwang In Kim (POSTECH, Republic of Korea), and Jaeho Lee (POSTECH, Republic of Korea)

Orals 2C 3D from Multiview and Sensors

Point Transformer V3: Simpler, Faster, Stronger 4840
Xiaoyang Wu (The Univeristy of Hong Kong, Hong Kong), Li Jiang (The Chinese University of Hong Kong, Shenzhen, China), Peng-Shuai Wang (Peking University, China), Zhijian Liu (Massachusetts Institute of Technology, USA), Xihui Liu (The Univeristy of Hong Kong, Hong Kong), Yu Qiao (Shanghai AI Laboratory, China), Wanli Ouyang (Shanghai AI Laboratory, China), Tong He (Shanghai AI Laboratory, China), and Hengshuang Zhao (The Univeristy of Hong Kong, Hong Kong)

Matching 2D Images in 3D: Metric Relative Pose from Metric Correspondences 4852
Axel Barroso-Laguna (Niantic), Sowmya Munukutla (Niantic), Victor Adrian Prisacariu (Niantic), and Eric Brachmann (Niantic)

Seeing the World through Your Eyes 4864
Hadi Alzayer (University of Maryland, USA), Kevin Zhang (University of Maryland, USA), Brandon Feng (University of Maryland, USA), Christopher A. Metzler (University of Maryland, USA), and Jia-Bin Huang (University of Maryland, USA)

Tri-Perspective View Decomposition for Geometry-Aware Depth Completion 4874
Zhiqiang Yan (Nanjing University of Science and Technology, China), Yuankai Lin (Huazhong University of Science and Technology), Kun Wang (Nanjing University of Science and Technology), Yupeng Zheng (Chinese Academy of Sciences), Yufei Wang (Northwestern Polytechnical University), Zhenyu Zhang (Nanjing University of Science and Technology), Jun Li (Nanjing University of Science and Technology), and Jian Yang (Nanjing University of Science and Technology)

Steerers: A Framework for Rotation Equivariant Keypoint Descriptors 4885
Georg Bökman (Chalmers University of Technology), Johan Edstedt (Linköping University), Michael Felsberg (Linköping University), and Fredrik Kahl (Chalmers University of Technology)

Poster Session 2

VP3D: Unleashing 2D Visual Prompt for Text-to-3D Generation 4896
Yang Chen (HiDream.ai Inc.), Yingwei Pan (HiDream.ai Inc.), Haibo Yang (Fudan University, China), Ting Yao (HiDream.ai Inc.), and Tao Mei (HiDream.ai Inc.)

Entangled View-Epipolar Information Aggregation for Generalizable Neural Radiance Fields 4906
Zhiyuan Min (Zhejiang University), Yawei Luo (Zhejiang University), Wei Yang (Huazhong University of Science and Technology), Yuesong Wang (Huazhong University of Science and Technology), and Yi Yang (Zhejiang University)

GroupContrast: Semantic-aware Self-supervised Representation Learning for 3D Understanding. 4917
Chengyao Wang (The Chinese University of Hong Kong, China), Li Jiang (The Chinese University of Hong Kong (Shenzhen), China), Xiaoyang Wu (University of Hong Kong, China), Zhuotao Tian (Harbin Institute of Technology (Shenzhen), China), Bohao Peng (The Chinese University of Hong Kong, China), Hengshuang Zhao (University of Hong Kong, China), and Jiaya Jia (The Chinese University of Hong Kong, China)

iToF-flow-based High Frame Rate Depth Imaging 4929
Yu Meng (Nanjing University, China), Zhou Xue (Li Auto, China), Xu Chang (Bytedance Inc, China), Xuemei Hu (Nanjing University, China), and Tao Yue (Nanjing University, China)

Generalizable Novel-View Synthesis using a Stereo Camera 4939
Haechan Lee (POSTECH, South Korea), Wonjoon Jin (POSTECH, South Korea), Seung-Hwan Baek (POSTECH, South Korea), and Sunghyun Cho (POSTECH, South Korea)

EfficientDreamer: High-Fidelity and Robust 3D Creation via Orthogonal-view Diffusion Priors 4949
Zhipeng Hu (NetEase Fuxi AI Lab), Minda Zhao (NetEase Fuxi AI Lab), Chaoyi Zhao (NetEase Fuxi AI Lab), Xinyue Liang (NetEase Fuxi AI Lab), Lincheng Li (NetEase Fuxi AI Lab), Zeng Zhao (NetEase Fuxi AI Lab), Changjie Fan (NetEase Fuxi AI Lab), Xiaowei Zhou (State Key Lab of CAD&CG, Zhejiang University), and Xin Yu (University of Queensland)

Leveraging Camera Triplets for Efficient and Accurate Structure-from-Motion 4959
Lalit Manam (Indian Institute of Science, India) and Venu Madhav Govindu (Indian Institute of Science, India)

LAENeRF: Local Appearance Editing for Neural Radiance Fields 4969
Lukas Radl (Graz University of Technology, Austria), Michael Steiner (Graz University of Technology, Austria), Andreas Kurz (Graz University of Technology, Austria), and Markus Steinberger (Graz University of Technology, Austria)

SuperPrimitive: Scene Reconstruction at a Primitive Level 4979
Kirill Mazur (Imperial College London, UK), Gwangbin Bae (Imperial College London, UK), and Andrew J. Davison (Imperial College London, UK)

Revisiting Sampson Approximations for Geometric Estimation Problems 4990
Felix Rydell (KTH Royal Institute of Technology, Sweden), Angélica Torres (Max Planck Institute for Mathematics in the Sciences, Germany), and Viktor Larsson (Lund University, Sweden)

Interactive3D: Create What You Want by Interactive 3D Generation 4999
Shaocong Dong (Hong Kong University of Science and Technology, China), Lihe Ding (The Chinese University of Hong Kong, China; Shanghai AI Laboratory, China), Zhanpeng Huang (SenseTime Research, China), Zibin Wang (SenseTime Research, China), Tianfan Xue (The Chinese University of Hong Kong, China), and Dan Xu (Hong Kong University of Science and Technology, China)

Multiplane Prior Guided Few-Shot Aerial Scene Rendering 5009
Zihan Gao (Xidian University, China), Licheng Jiao (Xidian University, China), Lingling Li (Xidian University, China), Xu Liu (Xidian University, China), Fang Liu (Xidian University, China), Puhua Chen (Xidian University, China), and Yuwei Guo (Xidian University, China)

3DGS-Avatar: Animatable Avatars via Deformable 3D Gaussian Splatting 5020
Zhiyin Qian (ETH Zürich, Switzerland), Shaofei Wang (ETH Zürich, Switzerland; University of Tübingen, Germany; Tübingen AI Center, Germany), Marko Mihajlovic (ETH Zürich, Switzerland), Andreas Geiger (University of Tübingen, Germany; Tübingen AI Center, Germany), and Siyu Tang (ETH Zürich, Switzerland)

DaReNeRF: Direction-aware Representation for Dynamic Scenes 5031
Ange Lou (Vanderbilt University, USA), Benjamin Planche (United Imaging Intelligence, USA), Zhongpai Gao (United Imaging Intelligence, USA), Yamin Li (Vanderbilt University, USA), Tianyu Luan (University of Buffalo, USA), Hao Ding (Johns Hopkins University, USA), Terrence Chen (United Imaging Intelligence, USA), Jack Noble (Vanderbilt University, USA), and Ziyan Wu (United Imaging Intelligence, USA)

ViewDiff: 3D-Consistent Image Generation with Text-to-Image Models 5043
Lukas Höllein (Technical University of Munich, Germany), Aljaž Božič (Meta, Switzerland), Norman Müller (Meta, Switzerland), David Novotny (Meta, England), Hung-Yu Tseng (Meta, USA), Christian Richardt (Meta, USA), Michael Zollhöfer (Meta, USA), and Matthias Nießner (Technical University of Munich, Germany)

LTM: Lightweight Textured Mesh Extraction and Refinement of Large Unbounded Scenes for Efficient Storage and Real-time Rendering 5053
Jaehoon Choi (University of Maryland), Rajvi Shah (Meta Reality Labs, USA), Qinbo Li (Meta Reality Labs, USA), Yipeng Wang (Meta Reality Labs, USA), Ayush Saraf (Meta Reality Labs, USA), Changil Kim (Meta Reality Labs, USA), Jia-Bin Huang (University of Maryland), Dinesh Manocha (University of Maryland), Suhib Alsisan (Meta Reality Labs, USA), and Johannes Kopf (Meta Reality Labs, USA)

Minimal Perspective Autocalibration 5064
Andrea Porfiri Dal Cin (Politecnico di Milano, Italy), Timothy Duff (University of Washington, USA), Luca Magri (Politecnico di Milano, Italy), and Tomas Pajdla (CIIRC CTU Prague, Czech Republic)

X-3D: Explicit 3D Structure Modeling for Point Cloud Recognition 5074
Shuofeng Sun (Beijing University of Posts and Telecommunications), Yongming Rao (Tencent), Jiwen Lu (Tsinghua University), and Haibin Yan (Beijing University of Posts and Telecommunications)

2S-UDF: A Novel Two-stage UDF Learning Method for Robust Non-watertight Model Reconstruction from Multi-view Images 5084
Junkai Deng (Institute of Software, Chinese Academy of Sciences, China), Fei Hou (Institute of Software, Chinese Academy of Sciences, China), Xuhui Chen (Institute of Software, Chinese Academy of Sciences, China), Wencheng Wang (Institute of Software, Chinese Academy of Sciences, China), and Ying He (Nanyang Technological University, Singapore)

UFORecon: Generalizable Sparse-View Surface Reconstruction from Arbitrary and Unfavorable Sets 5094
Youngju Na (KAIST, South Korea), Woo Jae Kim (KAIST, South Korea), Kyu Beom Han (KAIST, South Korea), Suhyeon Ha (KAIST, South Korea), and Sung-Eui Yoon (KAIST, South Korea)

GenN2N: Generative NeRF2NeRF Translation 5105
Xiangyue Liu (Hong Kong University of Science and Technology, China), Han Xue (Tsinghua University, China), Kunming Luo (Hong Kong University of Science and Technology, China), Ping Tan (Hong Kong University of Science and Technology, China), and Li Yi (Tsinghua University, China)

Text-to-3D Generation with Bidirectional Diffusion using both 2D and 3D priors 5115
Lihe Ding (The Chinese University of Hong Kong, China; Shanghai AI Laboratory, China), Shaocong Dong (Hong Kong University of Science and Technology, China), Zhanpeng Huang (SenseTime Research, China), Zibin Wang (SenseTime Research, China), Yiyuan Zhang (The Chinese University of Hong Kong, China), Kaixiong Gong (The Chinese University of Hong Kong, China), Dan Xu (Hong Kong University of Science and Technology, China), and Tianfan Xue (The Chinese University of Hong Kong, China)

Noisy One-point Homographies are Surprisingly Good 5125
Yaqing Ding (Czech Technical University in Prague, Czech Republic), Jonathan Astermark (Lund University, Sweden), Magnus Oskarsson (Lund University, Sweden), and Viktor Larsson (Lund University, Sweden)

Adaptive Multi-Modal Cross-Entropy Loss for Stereo Matching 5135
Peng Xu (Zhejiang University, China), Zhiyu Xiang (Zhejiang University, China), Chengyu Qiao (Zhejiang University, China), Jingyun Fu (Zhejiang University, China), and Tianyu Pu (Zhejiang University, China)

LiDAR4D: Dynamic Neural Fields for Novel Space-time View LiDAR Synthesis 5145
Zehan Zheng (Tongji University), Fan Lu (Tongji University), Weiyi Xue (Tongji University), Guang Chen (Tongji University), and Changjun Jiang (Tongji University)

NC-SDF: Enhancing Indoor Scene Reconstruction Using Neural SDFs with View-Dependent Normal Compensation 5155
Ziyi Chen (Zhejiang University, China), Xiaolong Wu (Amap Alibaba, China), and Yu Zhang (Zhejiang University, China)

VastGaussian: Vast 3D Gaussians for Large Scene Reconstruction 5166
Jiaqi Lin (Tsinghua University), Zhihao Li (Huawei Noah's Ark Lab), Xiao Tang (Huawei Noah's Ark Lab), Jianzhuang Liu (Shenzhen Institute of Advanced Technology), Shiyong Liu (Huawei Noah's Ark Lab), Jiayue Liu (Tsinghua University), Yangdi Lu (Huawei Noah's Ark Lab), Xiaofei Wu (Huawei Noah's Ark Lab), Songcen Xu (Huawei Noah's Ark Lab), Youliang Yan (Huawei Noah's Ark Lab), and Wenming Yang (Tsinghua University)

Language-driven Object Fusion into Neural Radiance Fields with Pose-Conditioned Dataset Updates 5176
Ka Chun Shum (Hong Kong University of Science and Technology), Jaeyeon Kim (Hong Kong University of Science and Technology), Binh-Son Hua (Trinity College Dublin; VinAI), Duc Thanh Nguyen (Deakin University), and Sai-Kit Yeung (Hong Kong University of Science and Technology)

SPU-PMD: Self-Supervised Point Cloud Upsampling via Progressive Mesh Deformation 5188
Yanzhe Liu (Dalian Maritime University, China), Rong Chen (Dalian Maritime University, China), Yushi Li (Xi'an Jiaotong-Liverpool University, China), Yixi Li (Dalian Maritime University, China), and Xuehou Tan (Tokai University, Japan)

Intrinsic Image Diffusion for Indoor Single-view Material Estimation 5198
Peter Kocsis (Technical University of Munich, Germany), Vincent Sitzmann (MIT EECS, USA), and Matthias Nießner (Technical Univesrity of Munich, Germany)

Learning Dynamic Tetrahedra for High-Quality Talking Head Synthesis 5209
Zicheng Zhang (University of Chinese Academy of Sciences), Ruobing Zheng (Ant Group), Bonan Li (University of Chinese Academy of Sciences), Congying Han (University of Chinese Academy of Sciences), Tianqi Li (Ant Group), Meng Wang (Ant Group), Tiande Guo (University of Chinese Academy of Sciences), Jingdong Chen (Ant Group), Ziwen Liu (University of Chinese Academy of Sciences), and Ming Yang (Ant Group)

Robust Self-calibration of Focal Lengths from the Fundamental Matrix 5220
Viktor Kocur (Comenius University in Bratislava, Slovakia; Czech Technical University in Prague, Czechia), Daniel Kyselica (Comenius University in Bratislava, Slovakia), and Zuzana Kukelova (Czech Technical University in Prague, Czechia)

RNb-NeuS: Reflectance and Normal-based Multi-View 3D Reconstruction 5230
Baptiste Brument (IRIT, France), Robin Bruneau (DIKU, Denmark), Yvain Quéau (GREYC, France), Jean Mélou (IRIT, France), François Bernard Lauze (DIKU, Denmark), Jean-Denis Durou (IRIT, France), and Lilian Calvet (Balgrist, Switzerland)

Neural 3D Strokes: Creating Stylized 3D Scenes with Vectorized 3D Strokes 5240
Hao-Bin Duan (State Key Laboratory of Virtual Reality Technology and Systems, SCSE, Beihang University), Miao Wang (State Key Laboratory of Virtual Reality Technology and Systems, SCSE, Beihang University; Zhongguanchun Laboratory), Yan-Xun Li (State Key Laboratory of Virtual Reality Technology and Systems, SCSE, Beihang University), and Yong-Liang Yang (University of Bath)

Unsupervised Template-assisted Point Cloud Shape Correspondence Network 5250
Jiacheng Deng (University of Science and Technology of China, China), Jiahao Lu (University of Science and Technology of China, China), and Tianzhu Zhang (University of Science and Technology of China, China; Deep Space Exploration Lab, China)

Efficient Detection of Long Consistent Cycles and its Application to Distributed Synchronization 5260
Shaohan Li (University of Minnesota, Twin Cities, USA), Yunpeng Shi (University of California, Davis, USA), and Gilad Lerman (University of Minnesota, Twin Cities, USA)

AirPlanes: Accurate Plane Estimation via 3D-Consistent Embeddings 5270
Jamie Watson (Niantic, UK; UCL, UK), Filippo Aleotti (Niantic, UK), Mohamed Sayed (Niantic, UK), Zawar Qureshi (Niantic, UK), Oisin Mac Aodha (University of Edinburgh, UK), Gabriel Brostow (Niantic, UK; UCL, UK), Michael Firman (Niantic, UK), and Sara Vicente (Niantic, UK)

Accurate Training Data for Occupancy Map Prediction in Automated Driving Using Evidence Theory 5281
Jonas Kälble (Bosch Center for Artificial Intelligence, Germany), Sascha Wirges (Bosch Center for Artificial Intelligence, Germany), Maxim Tatarchenko (Bosch Center for Artificial Intelligence, Germany), and Eddy Ilg (Saarland University, Germany)

Continuous Pose for Monocular Cameras in Neural Implicit Representation 5291
Qi Ma (Computer Vision Lab, ETH Zurich, Switzerland, INSAIT, Sofia University, Bulgaria), Danda Pani Paudel (INSAIT, Sofia University Bulgaria), Ajad Chhatkuli (Computer Vision Lab, ETH Zurich, Switzerland), and Luc Van Gool (Computer Vision Lab, ETH Zurich, Switzerland, INSAIT, Sofia University Bulgaria)

Towards 3D Vision with Low-Cost Single-Photon Cameras 5302
Fangzhou Mu (University of Wisconsin - Madison), Carter Sifferman (University of Wisconsin - Madison), Sacha Jungerman (University of Wisconsin - Madison), Yiquan Li (University of Wisconsin - Madison), Mark Han (University of Wisconsin - Madison), Michael Gleicher (University of Wisconsin - Madison), Mohit Gupta (University of Wisconsin - Madison), and Yin Li (University of Wisconsin - Madison)

Inlier Confidence Calibration for Point Cloud Registration 5312
Yongzhe Yuan (MoE Key Lab of Collaborative Intelligence Systems, Xidian University), Yue Wu (MoE Key Lab of Collaborative Intelligence Systems, Xidian University), Xiaolong Fan (MoE Key Lab of Collaborative Intelligence Systems, Xidian University), Maoguo Gong (MoE Key Lab of Collaborative Intelligence Systems, Xidian University), Qiguang Miao (MoE Key Lab of Collaborative Intelligence Systems, Xidian University), and Wenping Ma (Xidian University)

GaussianShader: 3D Gaussian Splatting with Shading Functions for Reflective Surfaces 5322
Yingwenqi Jiang (Shanghaitech University, China), Jiadong Tu (ShanghaiTech University, China), Yuan Liu (The University of Hong Kong), Xifeng Gao (Tencent America), Xiaoxiao Long (The University of Hong Kong), Wenping Wang (Texas A&M University), and Yuexin Ma (ShanghaiTech University, China)

Language Embedded 3D Gaussians for Open-Vocabulary Scene Understanding 5333
Jin-Chuan Shi (State Key Laboratory of Virtual Reality Technology and Systems, SCSE, Beihang University), Miao Wang (State Key Laboratory of Virtual Reality Technology and Systems, SCSE, Beihang University; Zhongguancun Laboratory), Hao-Bin Duan (State Key Laboratory of Virtual Reality Technology and Systems, SCSE, Beihang University), and Shao-Hua Guan (State Key Laboratory of Virtual Reality Technology and Systems, SCSE, Beihang University)

MVIP-NeRF: Multi-view 3D Inpainting on NeRF Scenes via Diffusion Prior 5344
Honghua Chen (Nanyang Technological University, Singapore), Chen Change Loy (Nanyang Technological University, Singapore), and Xingang Pan (Nanyang Technological University, Singapore)

SuGaR: Surface-Aligned Gaussian Splatting for Efficient 3D Mesh Reconstruction and High-Quality Mesh Rendering 5354
Antoine Guédon (Ecole des Ponts ParisTech, France) and Vincent Lepetit (Ecole des Ponts ParisTech, France)

DreamControl: Control-Based Text-to-3D Generation with 3D Self-Prior 5364
Tianyu Huang (Harbin Institute of Technology, China; City University of Hong Kong, Hong Kong S.A.R.), Yihan Zeng (Huawei Noah's Ark Lab, China), Zhilu Zhang (Harbin Institute of Technology, China), Wan Xu (Harbin Institute of Technology, China), Hang Xu (Huawei Noah's Ark Lab, China), Songcen Xu (Huawei Noah's Ark Lab, China), Rynson W.H. Lau (City University of Hong Kong, Hong Kong S.A.R.), and Wangmeng Zuo (Harbin Institute of Technology, China)

VAREN: Very Accurate and Realistic Equine Network 5374
Silvia Zuffi (IMATI-CNR, Milan, Italy), Ylva Mellbin (Swedish University of Agricultural Sciences, Sweden), Ci Li (KTH Royal Institute of Technology, Sweden), Markus Hoeschle (Max Planck Institute for Intelligent Systems, Tübingen, Germany), Hedvig Kjellström (KTH Royal Institute of Technology, Sweden), Senya Polikovsky (Max Planck Institute for Intelligent Systems, Tübingen, Germany), Elin Hernlund (Swedish University of Agricultural Sciences, Sweden), and Michael J. Black (Max Planck Institute for Intelligent Systems, Tübingen, Germany)

REACTO: Reconstructing Articulated Objects from a Single Video 5384
*Chaoyue Song (Nanyang Technological University, Singapore; A*STAR, Singapore), Jiacheng Wei (Nanyang Technological University, Singapore), Chuan Sheng Foo (A*STAR, Singapore), Guosheng Lin (Nanyang Technological University, Singapore), and Fayao Liu (A*STAR, Singapore)*

DITTO: Dual and Integrated Latent Topologies for Implicit 3D Reconstruction 5396
Jaehyeok Shim (Ulsan National Institute of Science and Technology, Republic of Korea) and Kyungdon Joo (Ulsan National Institute of Science and Technology, Republic of Korea)

ICON: Incremental CONfidence for Joint Pose and Radiance Field Optimization 5406
Weiyao Wang (FAIR, Meta), Pierre Gleize (FAIR, Meta), Hao Tang (FAIR, Meta), Xingyu Chen (FAIR, Meta), Kevin J Liang (FAIR, Meta), and Matt Feiszli (FAIR, Meta)

Local-consistent Transformation Learning for Rotation-invariant Point Cloud Analysis 5418
Yiyang Chen (South China University of Technology), Lunhao Duan (Wuhan University), Shanshan Zhao (The University of Sydney), Changxing Ding (South China University of Technology; Pazhou Lab), and Dacheng Tao (Nanyang Technological University)

PaReNeRF: Toward Fast Large-scale Dynamic NeRF with Patch-based Reference 5428
Xiao Tang (Samsung R&D Institute China Xi'an (SRCX), China), Min Yang (Samsung R&D Institute China Xi'an (SRCX), China), Penghui Sun (Samsung R&D Institute China Xi'an (SRCX), China), Hui Li (Samsung R&D Institute China Xi'an (SRCX), China), Yuchao Dai (Northwestern Polytechnical University, China), Feng Zhu (Samsung R&D Institute China Xi'an (SRCX), China), and Hojae Lee (Samsung R&D Institute China Xi'an (SRCX), China)

Fitting Flats to Flats 5439
Gabriel Dogadov (TU Berlin), Ugo Finnendahl (TU Berlin), and Marc Alexa (TU Berlin)

ANIM: Accurate Neural Implicit Model for Human Reconstruction from a single RGB-D Image .. 5448
Marco Pesavento (University of Surrey, UK), Yuanlu Xu (Meta Reality Labs, USA), Nikolaos Sarafianos (Meta Reality Labs, USA), Robert Maier (Meta Reality Labs, Switzerland), Ziyan Wang (Meta Reality Labs, USA), Chun-Han Yao (UC Merced, USA), Marco Volino (University of Surrey, UK), Edmond Boyer (Meta Reality Labs, USA), Adrian Hilton (University of Surrey, UK), and Tony Tung (Meta Reality Labs, USA)

Neural Markov Random Field for Stereo Matching 5459
Tongfan Guan (The Chinese University of Hong Kong, China), Chen Wang (University at Buffalo, USA), and Yun-Hui Liu (The Chinese University of Hong Kong, China)

Improving Physics-Augmented Continuum Neural Radiance Field-Based Geometry-Agnostic System Identification with Lagrangian Particle Optimization 5470
Takuhiro Kaneko (NTT Corporation, Japan)

DiffusionAvatars: Deferred Diffusion for High-fidelity 3D Head Avatars 5481
Tobias Kirschstein (Technical University of Munich, Germany), Simon Giebenhain (Technical University of Munich, Germany), and Matthias Nießner (Technical University of Munich, Germany)

ViT-CoMer: Vision Transformer with Convolutional Multi-scale Feature Interaction for Dense Predictions 5493
Chunlong Xia (Baidu, China), Xinliang Wang (Baidu, China), Feng Lv (Baidu, China), Xin Hao (Baidu, China), and Yifeng Shi (Baidu, China)

Pose-Transformed Equivariant Network for 3D Point Trajectory Prediction 5503
Ruixuan Yu (Shandong University, Weihai) and Jian Sun (Xi'an Jiaotong University)

UniRepLKNet: A Universal Perception Large-Kernel ConvNet for Audio, Video, Point Cloud, Time-Series and Image Recognition 5513
Xiaohan Ding (Tencent AI Lab), Yiyuan Zhang (The Chinese University of Hong Kong), Yixiao Ge (Tencent AI Lab), Sijie Zhao (Tencent AI Lab), Lin Song (Tencent AI Lab), Xiangyu Yue (The Chinese University of Hong Kong), and Ying Shan (Tencent AI Lab)

KPConvX: Modernizing Kernel Point Convolution with Kernel Attention 5525
Hugues Thomas (Apple, USA), Yao-Hung Hubert Tsai (Apple, USA), Timothy D. Barfoot (University of Toronto, Canada), and Jian Zhang (Apple, USA)

Time-, Memory- and Parameter-Efficient Visual Adaptation 5536
Otniel-Bogdan Mercea (University of Tubingen), Alexey Gritsenko (Google), Cordelia Schmid (Google), and Anurag Arnab (Google)

Affine Equivariant Networks Based on Differential Invariants 5546
Yikang Li (Peking University, China), Yeqing Qiu (The Chinese University of Hong Kong, Shenzhen, China; Shenzhen Research Institute of Big Data, China), Yuxuan Chen (Peking University, China), Lingshen He (Peking University, China), and Zhouchen Lin (Peking University, China; Peng Cheng Laboratory, China)

PeLK: Parameter-efficient Large Kernel ConvNets with Peripheral Convolution 5557
Honghao Chen (Institute of Automation, Chinese Academy of Science), Xiangxiang Chu (Meituan), Yongjian Ren (Institute of Automation, Chinese Academy of Science), Xin Zhao (Institute of Automation, Chinese Academy of Science), and Kaiqi Huang (Institute of Automation, Chinese Academy of Science)

Making Vision Transformers Truly Shift-Equivariant 5568
Renan A. Rojas-Gomez (University of Illinois at Urbana-Champaign), Teck-Yian Lim (University of Illinois at Urbana-Champaign), Minh N. Do (University of Illinois at Urbana-Champaign), and Raymond A. Yeh (Purdue University)

Once for Both: Single Stage of Importance and Sparsity Search for Vision Transformer Compression 5578
Hancheng Ye (Fudan University; Shanghai Artificial Intelligence Laboratory), Chong Yu (Fudan University), Peng Ye (Fudan University), Renqiu Xia (Shanghai Artificial Intelligence Laboratory; Shanghai Jiao Tong University), Yansong Tang (Tsinghua University), Jiwen Lu (Tsinghua University), Tao Chen (Fudan University), and Bo Zhang (Shanghai Artificial Intelligence Laboratory)

Data-Free Quantization via Pseudo-label Filtering 5589
Chunxiao Fan (Hefei University of Technology, China), Ziqi Wang (Hefei University of Technology, China), Dan Guo (Hefei University of Technology, China), and Meng Wang (Hefei University of Technology, China)

FedHCA2: Towards Hetero-Client Federated Multi-Task Learning 5599
Yuxiang Lu (Shanghai Jiao Tong University, China), Suizhi Huang (Shanghai Jiao Tong University, China), Yuwen Yang (Shanghai Jiao Tong University, China), Shalayiding Sirejiding (Shanghai Jiao Tong University, China), Yue Ding (Shanghai Jiao Tong University, China), and Hongtao Lu (Shanghai Jiao Tong University, China)

SpikingResformer: Bridging ResNet and Vision Transformer in Spiking Neural Networks 5610
Xinyu Shi (Peking University, China), Zecheng Hao (Peking University, China), and Zhaofei Yu (Peking University, China)

TetraSphere: A Neural Descriptor for O(3)-Invariant Point Cloud Analysis 5620
Pavlo Melnyk (Linköping University, Sweden), Andreas Robinson (Linköping University, Sweden), Michael Felsberg (Linköping University, Sweden), and Mårten Wadenbäck (Linköping University, Sweden)

Friendly Sharpness-Aware Minimization 5631
Tao Li (Shanghai Jiao Tong University, China), Pan Zhou (Singapore Management University, Singapore), Zhengbao He (Shanghai Jiao Tong University, China), Xinwen Cheng (Shanghai Jiao Tong University, China), and Xiaolin Huang (Shanghai Jiao Tong University, China)

RMT: Retentive Networks Meet Vision Transformers 5641
Qihang Fan (Institute of Automation, Chinese Academy of Sciences, China), Huaibo Huang (Institute of Automation, Chinese Academy of Sciences, China), Mingrui Chen (Institute of Automation, Chinese Academy of Sciences, China), Hongmin Liu (University of Science and Technology Beijing, China), and Ran He (Institute of Automation, Chinese Academy of Sciences, China)

Efficient Deformable ConvNets: Rethinking Dynamic and Sparse Operator for Vision Applications 5652
Yuwen Xiong (University of Toronto; Shanghai AI Lboratory), Zhiqi Li (Nanjing University; Shanghai AI Laboratory), Yuntao Chen (CAIR, HKISI, CAS), Feng Wang (Tsinghua University), Xizhou Zhu (Tsinghua University), Jiapeng Luo (Sensetime Research), Wenhai Wang (The Chinese University of Hong Kong; Shanghai AI Laboratory), Tong Lu (Nanjing University), Hongsheng Li (The Chinese University of Hong Kong), Yu Qiao (Shanghai AI Laboratory), Lewei Lu (Sensetime Research), Jie Zhou (Tsinghua University), and Jifeng Dai (Tsinghua University; Shanghai AI Laboratory)

Boosting Order-Preserving and Transferability for Neural Architecture Search: a Joint Architecture Refined Search and Fine-tuning Approach 5662
Beichen Zhang (Shanghai Jiao Tong University), Xiaoxing Wang (Shanghai Jiao Tong University), Xiaohan Qin (Shanghai Jiao Tong University), and Junchi Yan (Shanghai Jiao Tong University)

InceptionNeXt: When Inception Meets ConvNeXt 5672
Weihao Yu (National University of Singapore, Singapore), Pan Zhou (Singapore Management University, Singapore), Shuicheng Yan (Skywork AI, Singapore), and Xinchao Wang (National University of Singapore, Singapore)

BiPer: Binary Neural Networks using a Periodic Function 5684
Edwin Vargas (Universidad Industrial de Santander), Claudia V. Correa (Universidad Industrial de Santander), Carlos Hinojosa (King Abdullah University of Science and Technology), and Henry Arguello (Universidad Industrial de Santander)

Rewrite the Stars 5694
Xu Ma (Northeastern University), Xiyang Dai (Microsoft), Yue Bai (Northeastern University), Yizhou Wang (Northeastern University), and Yun Fu (Northeastern University)

A&B BNN: Add&Bit-Operation-Only Hardware-Friendly Binary Neural Network 5704
Ruichen Ma (University of Electronic Science and Technology of China), Guanchao Qiao (University of Electronic Science and Technology of China), Yian Liu (University of Electronic Science and Technology of China), Liwei Meng (University of Electronic Science and Technology of China), Ning Ning (University of Electronic Science and Technology of China), Yang Liu (University of Electronic Science and Technology of China), and Shaogang Hu (University of Electronic Science and Technology of China)

Neural Clustering based Visual Representation Learning 5714
Guikun Chen (Zhejiang University), Xia Li (ETH Zurich), Yi Yang (Zhejiang University), and Wenguan Wang (Zhejiang University)

Building Optimal Neural Architectures using Interpretable Knowledge 5726
Keith G. Mills (University of Alberta, Canada), Fred X. Han (Huawei Technologies, Canada), Mohammad Salameh (Huawei Technologies, Canada), Shengyao Lu (University of Alberta, Canada), Chunhua Zhou (Huawei Kirin Solution, China), Jiao He (Huawei Kirin Solution, China), Fengyu Sun (Huawei Kirin Solution, China), and Di Niu (University of Alberta, Canada)

Towards More Accurate Diffusion Model Acceleration with A Timestep Tuner 5736
Mengfei Xia (Tsinghua University, China), Yujun Shen (Ant Group, China), Changsong Lei (Tsinghua University, China), Yu Zhou (Tsinghua University, China), Deli Zhao (Alibaba Group, China), Ran Yi (Shanghai Jiao Tong University, China), Wenping Wang (Texas A&M University, USA), and Yong-Jin Liu (Tsinghua University, China)

UniPTS: A Unified Framework for Proficient Post-Training Sparsity 5746
Jingjing Xie (Xiamen University, China), Yuxin Zhang (Xiamen University, China), Mingbao Lin (Tencent Youtu Lab, China), Liujuan Cao (Xiamen University, China), and Rongrong Ji (Xiamen University, China)

SHViT: Single-Head Vision Transformer with Memory Efficient Macro Design 5756
Seokju Yun (University of Seoul) and Youngmin Ro (University of Seoul)

Denoising Point Clouds in Latent Space via Graph Convolution and Invertible Neural Network… 5768
Aihua Mao (South China University of Technology), Biao Yan (South China University of Technology), Zijing Ma (South China University of Technology), and Ying He (Nanyang Technological University)

JointSQ: Joint Sparsification-Quantization for Distributed Learning 5778
Weiying Xie (XiDian University, China), Haowei Li (XiDian University, China), Jitao Ma (XiDian University, China), Yunsong Li (XiDian University, China), Jie Lei (University of Technology Sydney, Australia), Donglai Liu (XiDian University, China), and Leyuan Fang (Hunan University)

YolOOD: Utilizing Object Detection Concepts for Multi-Label Out-of-Distribution Detection 5788
Alon Zolfi (Ben-Gurion University of the Negev, Israel), Guy Amit (Ben-Gurion University of the Negev, Israel), Amit Baras (Ben-Gurion University of the Negev, Israel), Satoru Koda (Fujitsu Limited, Japan), Ikuya Morikawa (Fujitsu Limited, Japan), Yuval Elovici (Ben-Gurion University of the Negev, Israel), and Asaf Shabtai (Ben-Gurion University of the Negev, Israel)

RepAn: Enhanced Annealing through Re-parameterization 5798
Xiang Fei (Xiamen University, China), Xiawu Zheng (Xiamen University, China), Yan Wang (Samsara, USA), Fei Chao (Xiamen University, China), Chenglin Wu (DeepWisdom Inc, China), and Liujuan Cao (Xiamen University, China)

D^4: Dataset Distillation via Disentangled Diffusion Model 5809
Duo Su (University of Chinese Academy of Sciences, China), Junjie Hou (University of Chinese Academy of Sciences, China), Weizhi Gao (North Carolina State University, USA), Yingjie Tian (University of Chinese Academy of Sciences, China), and Bowen Tang (University of Chinese Academy of Sciences, China)

State Space Models for Event Cameras 5819
Nikola Zubic (University of Zurich), Mathias Gehrig (University of Zurich), and Davide Scaramuzza (University of Zurich)

Your Image is My Video: Reshaping the Receptive Field via Image-To-Video Differentiable AutoAugmentation and Fusion 5829
Sofia Casarin (Free University of Bozen-Bolzano, Italy), Cynthia I. Ugwu (Free University of Bozen-Bolzano, Italy), Sergio Escalera (Computer Vision Center, Spain; Universitat de Barcelona, Spain), and Oswald Lanz (Free University of Bozen-Bolzano, Italy)

Sparse Semi-DETR: Sparse Learnable Queries for Semi-Supervised Object Detection 5840
Tahira Shehzadi (German Research Center for AI (dfki.de), Germany), Khurram Azeem Hashmi (German Research Center for AI (dfki.de), Germany), Didier Stricker (German Research Center for AI (dfki.de), Germany), and Muhammad Zeshan Afzal (German Research Center for AI (dfki.de), Germany)

MAPSeg: Unified Unsupervised Domain Adaptation for Heterogeneous Medical Image Segmentation Based on 3D Masked Autoencoding and Pseudo-Labeling 5851
Xuzhe Zhang (Columbia University, USA), Yuhao Wu (Duke University, USA), Elsa Angelini (Télécom ParisTech, France), Ang Li (University of Maryland, College Park, USA), Jia Guo (Columbia University, USA), Jerod M. Rasmussen (University of California, Irvine, USA), Thomas G. O'Connor (University of Rochester, USA), Pathik D. Wadhwa (University of California, Irvine, USA), Andrea Parolin Jackowski (Østfold University College, Norway), Hai Li (Duke University, USA), Jonathan Posner (Duke University, USA), Andrew F. Laine (Columbia University, USA), and Yun Wang (Emory University, USA)

FedUV: Uniformity and Variance for Heterogeneous Federated Learning 5863
Ha Min Son (University of California, Davis), Moon-Hyun Kim (Hippo T&C; Sungkyunkwan University, Korea), Tai-Myoung Chung (Hippo T&C; Sungkyunkwan University, Korea), Chao Huang (University of California, Davis), and Xin Liu (University of California, Davis)

Pick-or-Mix: Dynamic Channel Sampling for ConvNets 5873
Ashish Kumar (Indian Institute of Technology, Kanpur, India), Daneul Kim (Seoul National University, South Korea), Jaesik Park (Seoul National University, South Korea), and Laxmidhar Behera (Indian Institute of Technolgy, Kanpur, India)

Sheared Backpropagation for Fine-tuning Foundation Models 5883
Zhiyuan Yu (University of Science and Technology of China, China), Li Shen (JD Explore Academy, China), Liang Ding (JD Explore Academy, China), Xinmei Tian (University of Science and Technology of China, China), Yixin Chen (Washington University in St. Louis, USA), and Dacheng Tao (Nanyang Technological University, Singapore)

AZ-NAS: Assembling Zero-Cost Proxies for Network Architecture Search 5893
Junghyup Lee (Yonsei University, Republic of Korea) and Bumsub Ham (Yonsei University, Republic of Korea; Korea Institute of Science and Technology (KIST), Republic of Korea)

MRFP: Learning Generalizable Semantic Segmentation from Sim-2-Real with Multi-Resolution Feature Perturbation 5904
Sumanth Udupa (Indian Institute of Science), Prajwal Gurunath (Indian Institute of Science), Aniruddh Sikdar (Indian Institute of Science), and Suresh Sundaram (Indian Institute of Science)

Training-Free Pretrained Model Merging 5915
Zhengqi Xu (Zhejiang University), Ke Yuan (Zhejiang University), Huiqiong Wang (Ningbo Innovation Center, Zhejiang University), Yong Wang (State Grid Shandong Electric Power Company), Mingli Song (Zhejiang University), and Jie Song (Zhejiang University)

Training Generative Image Super-Resolution Models by Wavelet-Domain Losses Enables Better Control of Artifacts 5926
Cansu Korkmaz (Koc University, Turkey), A. Murat Tekalp (Koc University, Turkey), and Zafer Dogan (Koc University, Turkey)

IReNe: Instant Recoloring of Neural Radiance Fields 5937
Alessio Mazzucchelli (Arquimea Research Center, Spain - Universitat Politècnica de Catalunya, Spain), Adrian Garcia-Garcia (Arquimea Research Center), Elena Garces (SEDDI, Spain - Universidad Rey Juan Carlos, Spain), Fernando Rivas-Manzaneque (Volinga AI, Spain - Universidad Politècnica de Madrid, Spain), Francesc Moreno-Noguer (Institut de Robòtica i Informática Industrial, Spain), and Adrian Penate-Sanchez (Universidad de las Palmas de Gran Canaria, Spain)

AdaShift: Learning Discriminative Self-Gated Neural Feature Activation With an Adaptive Shift Factor 5947
Sudong Cai (Kyoto University, Japan)

Kernel Adaptive Convolution for Scene Text Detection via Distance Map Prediction 5957
Jinzhi Zheng (Institute of Software Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), Heng Fan (University of North Texas, USA), and Libo Zhang (Institute of Software Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China)

Towards Accurate and Robust Architectures via Neural Architecture Search 5967
Yuwei Ou (Sichuan University, China), Yuqi Feng (Sichuan University, China), and Yanan Sun (Sichuan University, China)

PDF: A Probability-Driven Framework for Open World 3D Point Cloud Semantic Segmentation .. 5977
Jinfeng Xu (Huazhong University of Science and Technology, China), Siyuan Yang (Huazhong University of Science and Technology, China), Xianzhi Li (Huazhong University of Science and Technology, China), Yuan Tang (Huazhong University of Science and Technology, China), Yixue Hao (Huazhong University of Science and Technology, China), Long Hu (Huazhong University of Science and Technology, China), and Min Chen (Huazhong University of Science and Technology, China)

Permutation Equivariance of Transformers and Its Applications 5987
Hengyuan Xu (Shanghai Jiao Tong University, China), Liyao Xiang (Shanghai Jiao Tong University, China), Hangyu Ye (Shanghai Jiao Tong University, China), Dixi Yao (University of Toronto, Canada), Pengzhi Chu (Shanghai Jiao Tong University, China), and Baochun Li (University of Toronto, Canada)

MedBN: Robust Test-Time Adaptation against Malicious Test Samples 5997
Hyejin Park (Pohang University of Science and Technology (POSTECH), South Korea), Jeongyeon Hwang (Pohang University of Science and Technology (POSTECH), South Korea), Sunung Mun (Pohang University of Science and Technology (POSTECH), South Korea), Sangdon Park (Pohang University of Science and Technology (POSTECH), South Korea), and Jungseul Ok (Pohang University of Science and Technology (POSTECH), South Korea)

Small Scale Data-Free Knowledge Distillation 6008
He Liu (China Telecom Cloud Technology Co., Ltd., Tsinghua University, China), Yikai Wang (Tsinghua University, China), Huaping Liu (Tsinghua University, China), Fuchun Sun (Tsinghua University, China), and Anbang Yao (Intel Labs China, China)

Identifying Important Group of Pixels using Interactions 6017
Kosuke Sumiyasu (Chiba University, Japan), Kazuhiko Kawamoto (Chiba University, Japan), and Hiroshi Kera (Chiba University, Japan)

Efficiently Assemble Normalization Layers and Regularization for Federated Domain Generalization 6027
Khiem Le (University of Notre Dame, USA), Long Ho (VinUniversity, Vietnam), Cuong Do (VinUniversity, Vietnam), Danh Le-Phuoc (Technical University Berlin, Germany), and Kok-Seng Wong (VinUniversity, Vietnam)

OrthCaps: An Orthogonal CapsNet with Sparse Attention Routing and Pruning 6037
Xinyu Geng (Harbin Institute of Technology Shenzhen, China), Jiaming Wang (Harbin Institute of Technology Shenzhen, China), Jiawei Gong (Harbin Institute of Technology Shenzhen, China), Yuerong Xue (Harbin Institute of Technology Shenzhen, China), Jun Xu (Harbin Institute of Technology Shenzhen, China), Fanglin Chen (Harbin Institute of Technology Shenzhen, China), and Xiaolin Huang (Shanghai Jiao Tong University)

Mean-Shift Feature Transformer 6047
Takumi Kobayashi (National Institute of Advanced Industrial Science and Technology, Japan)

You Only Need Less Attention at Each Stage in Vision Transformers 6057
Shuoxi Zhang (Huazhong University of Science and Technology), Hanpeng Liu (Huazhong University of Science and Technology), Stephen Lin (Microsoft Research Asia), and Kun He (Huazhong University of Science and Technology)

HEAL-SWIN: A Vision Transformer On The Sphere 6067
Oscar Carlsson (Chalmers University of Technology and University of Gothenburg, Sweden), Jan E. Gerken (Chalmers University of Technology and University of Gothenburg, Sweden), Hampus Linander (Chalmers University of Technology and University of Gothenburg, Sweden), Heiner Spieß (Technical University Berlin, Germany), Fredrik Ohlsson (Umeå University, Sweden), Christoffer Petersson (Zenseact and Chalmers University of Technology and University of Gothenburg, Sweden), and Daniel Persson (Chalmers University of Technology and University of Gothenburg, Sweden)

NC-TTT: A Noise Constrastive Approach for Test-Time Training 6078
David Osowiechi (Ecole de Technologie Superieure), Gustavo A. Vargas Hakim (Ecole de Technologie Superieure), Mehrdad Noori (Ecole de Technologie Superieure), Milad Cheraghalikhani (Ecole de Technologie Superieure), Ali Bahri (Ecole de Technologie Superieure), Moslem Yazdanpanah (Ecole de Technologie Superieure), Ismail Ben Ayed (Ecole de Technologie Superieure), and Christian Desrosiers (Ecole de Technologie Superieure)

Unlocking the Potential of Prompt-Tuning in Bridging Generalized and Personalized Federated Learning 6087
Wenlong Deng (University of British Columbia, Canada), Christos Thrampoulidis (University of British Columbia, Canada), and Xiaoxiao Li (University of British Columbia, Canada)

MR-VNet: Media Restoration using Volterra Networks 6098
Siddharth Roheda (Samsung Research Institute), Amit Unde (Samsung Research Institute), and Loay Rashid (Samsung Research Institute)

Multimodal Pathway: Improve Transformers with Irrelevant Data from Other Modalities 6108
Yiyuan Zhang (The Chinese University of Hong Kong), Xiaohan Ding (Tencent AI Lab), Kaixiong Gong (The Chinese University of Hong Kong), Yixiao Ge (Tencent AI Lab), Ying Shan (Tencent AI Lab), and Xiangyu Yue (The Chinese University of Hong Kong)

GreedyViG: Dynamic Axial Graph Construction for Efficient Vision GNNs 6118
Mustafa Munir (The University of Texas at Austin, USA), William Avery (The University of Texas at Austin, USA), Md Mostafijur Rahman (The University of Texas at Austin, USA), and Radu Marculescu (The University of Texas at Austin, USA)

FlowerFormer: Empowering Neural Architecture Encoding using a Flow-aware Graph Transformer.... 6128
Dongyeong Hwang (Kim Jaechul Graduate School of AI, KAIST, Seoul, Republic of Korea), Hyunju Kim (Kim Jaechul Graduate School of AI, KAIST, Seoul, Republic of Korea), Sunwoo Kim (Kim Jaechul Graduate School of AI, KAIST, Seoul, Republic of Korea), and Kijung Shin (Kim Jaechul Graduate School of AI, KAIST, Seoul, Republic of Korea)

Mixed-Precision Quantization for Federated Learning on Resource-Constrained Heterogeneous Devices 6138
Huancheng Chen (University of Texas at Austin, USA) and Haris Vikalo (University of Texas at Austin, USA)

Wired Perspectives: Multi-View Wire Art Embraces Generative AI 6149
Zhiyu Qu (SketchX, CVSSP, University of Surrey), Lan Yang (Beijing University of Posts and Telecommunications), Honggang Zhang (Beijing University of Posts and Telecommunications), Tao Xiang (SketchX, CVSSP, University of Surrey), Kaiyue Pang (SketchX, CVSSP, University of Surrey), and Yi-Zhe Song (SketchX, CVSSP, University of Surrey)

DemoFusion: Democratising High-Resolution Image Generation With No $$$ 6159
Ruoyi Du (Beijing University of Posts and Telecommunications), Dongliang Chang (Tsinghua University), Timothy Hospedales (University of Edinburgh), Yi-Zhe Song (University of Surrey), and Zhanyu Ma (Beijing University of Posts and Telecommunications)

DiffPerformer: Iterative Learning of Consistent Latent Guidance for Diffusion-based Human Video Generation 6169
Chenyang Wang (Harbin Institute of Technology, China), Zerong Zheng (NNKosmos Technology, China), Tao Yu (Tsinghua University, China), Xiaoqian Lv (Harbin Institute of Technology, China), Bineng Zhong (Guangxi Normal University, China), Shengping Zhang (Harbin Institute of Technology, China), and Liqiang Nie (Harbin Institute of Technology, China)

InteractDiffusion: Interaction Control in Text-to-Image Diffusion Models 6180
Jiun Tian Hoe (Nanyang Technological University, Singapore), Xudong Jiang (Nanyang Technological University, Singapore), Chee Seng Chan (Universiti Malaya, Malaysia), Yap-Peng Tan (Nanyang Technological University, Singapore), and Weipeng Hu (Nanyang Technological University, Singapore)

Intelligent Grimm - Open-ended Visual Storytelling via Latent Diffusion Models 6190
Chang Liu (Shanghai Jiao Tong University, China; Shanghai AI lab, China), Haoning Wu (Shanghai Jiao Tong University, China), Yujie Zhong (Meituan Inc, China), Xiaoyun Zhang (Shanghai Jiao Tong University, China), Yanfeng Wang (Shanghai Jiao Tong University, China; Shanghai AI lab, China), and Weidi Xie (Shanghai Jiao Tong University, China; Shanghai AI lab, China)

ControlRoom3D: Room Generation using Semantic Proxy Rooms 6201
Jonas Schult (RWTH Aachen University, Germany), Sam Tsai (Meta GenAI, USA), Lukas Höllein (Technical University of Munich, Germany), Bichen Wu (Meta GenAI, USA), Jialiang Wang (Meta GenAI, USA), Chih-Yao Ma (Meta GenAI, USA), Kunpeng Li (Meta GenAI, USA), Xiaofang Wang (Meta GenAI, USA), Felix Wimbauer (Technical University of Munich, Germany), Zijian He (Meta GenAI, USA), Peizhao Zhang (Meta GenAI, USA), Bastian Leibe (RWTH Aachen University, Germany), Peter Vajda (Meta GenAI, USA), and Ji Hou (Meta GenAI, USA)

Cache Me if You Can: Accelerating Diffusion Models through Block Caching 6211
Felix Wimbauer (Technical University of Munich, Germany), Bichen Wu (Meta GenAI), Edgar Schoenfeld (Meta GenAI), Xiaoliang Dai (Meta GenAI), Ji Hou (Meta GenAI), Zijian He (Meta GenAI), Artsiom Sanakoyeu (Meta GenAI), Peizhao Zhang (Meta GenAI), Sam Tsai (Meta GenAI), Jonas Kohler (Meta GenAI), Christian Rupprecht (University of Oxford), Daniel Cremers (Technical University of Munich), Peter Vajda (Meta GenAI), and Jialiang Wang (Meta GenAI)

Real-time 3D-aware Portrait Video Relighting 6221
Ziqi Cai (Institute of Computing Technology, Chinese Academy of Sciences, China; Beijing Jiaotong University, China), Kaiwen Jiang (University of California San Diego, USA), Shu-Yu Chen (Institute of Computing Technology, Chinese Academy of Sciences, China), Yu-Kun Lai (Cardiff University, UK), Hongbo Fu (City University of Hong Kong, China; The Hong Kong University of Science and Technology, China), Boxin Shi (Peking University, China), and Lin Gao (Institute of Computing Technology, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China)

InstanceDiffusion: Instance-level Control for Image Generation 6232
Xudong Wang (UC Berkeley, USA), Trevor Darrell (UC Berkeley, USA), Sai Saketh Rambhatla (Meta AI, USA), Rohit Girdhar (Meta AI, USA), and Ishan Misra (Meta AI, USA)

Make-It-Vivid: Dressing Your Animatable Biped Cartoon Characters from Text 6243
Junshu Tang (Shanghai Jiao Tong University), Yanhong Zeng (Shanghai AI Lab), Ke Fan (Shanghai Jiao Tong University), Xuheng Wang (Tsinghua University), Bo Dai (Shanghai AI Lab), Kai Chen (Shanghai AI Lab), and Lizhuang Ma (Shanghai Jiao Tong University)

ZONE: Zero-Shot Instruction-Guided Local Editing 6254
Shanglin Li (Beihang University, China), Bohan Zeng (Beihang University, China), Yutang Feng (Beihang University, China), Sicheng Gao (Beihang University, China), Xiuhui Liu (Beihang University, China), Jiaming Liu (Xiaohongshu Inc, China), Lin Li (Xiaohongshu Inc, China), Xu Tang (Xiaohongshu Inc, China), Yao Hu (Xiaohongshu Inc, China), Jianzhuang Liu (Shenzhen Institute of Advanced Technology, China), and Baochang Zhang (Beihang University, China)

Don't Drop Your Samples! Coherence-Aware Training Benefits Conditional Diffusion 6264
Nicolas Dufour (LIGM, École des Ponts, Univ Gustave Eiffel, CNRS, Marne-la-Vallée, France), Victor Besnier (Valeo ai, Prague), Vicky Kalogeiton (LIX, CNRS, École Polytechnique, IP Paris), and David Picard (LIGM, École des Ponts, Univ Gustave Eiffel, CNRS, Marne-la-Vallée, France)

Generating Illustrated Instructions 6274
Sachit Menon (Meta), Ishan Misra (Meta), and Rohit Girdhar (Meta)

SpikeNeRF: Learning Neural Radiance Fields from Continuous Spike Stream 6285
Lin Zhu (Beijing Institute of Technology), Kangmin Jia (Beijing Institute of Technology), Yifan Zhao (Beihang University), Yunshan Qi (Beihang University), Lizhi Wang (Beijing Institute of Technology), and Hua Huang (Beijing Normal University)

Dancing with Still Images: Video Distillation via Static-Dynamic Disentanglement 6296
Ziyu Wang (Shanghai Jiao Tong University, China), Yue Xu (Shanghai Jiao Tong University, China), Cewu Lu (Shanghai Jiao Tong University, China), and Yong-Lu Li (Shanghai Jiao Tong University, China)

UniGS: Unified Representation for Image Generation and Segmentation 6305
Lu Qi (The University of California Merced, USA), Lehan Yang (The University of Sydney, Australia), Weidong Guo (Tencent, China), Yu Xu (Tencent, China), Bo Du (Wuhan University), Varun Jampani (Stability AI, USA), and Ming-Hsuan Yang (The University of California Merced, USA)

Adversarial Text to Continuous Image Generation 6316
Kilichbek Haydarov (King Abdullah University of Science and Technology), Aashiq Muhamed (King Abdullah University of Science and Technology), Xiaoqian Shen (King Abdullah University of Science and Technology), Jovana Lazarevic (King Abdullah University of Science and Technology), Ivan Skorokhodov (King Abdullah University of Science and Technology), Chamuditha Jayanga Galappaththige (King Abdullah University of Science and Technology), and Mohamed Elhoseiny (King Abdullah University of Science and Technology)

Self-correcting LLM-controlled Diffusion Models 6327
Tsung-Han Wu (University of California, Berkeley), Long Lian (University of California, Berkeley), Joseph E. Gonzalez (University of California, Berkeley), Boyi Li (University of California, Berkeley), and Trevor Darrell (University of California, Berkeley)

TiNO-Edit: Timestep and Noise Optimization for Robust Diffusion-Based Image Editing 6337
Sherry X Chen (University of California, Santa Barbara), Yaron Vaxman (Cloudinary), Elad Ben Baruch (Cloudinary), David Asulin (Cloudinary), Aviad Moreshet (Cloudinary), Kuo-Chin Lien (Layer AI), Misha Sra (University of California, Santa Barbara), and Pradeep Sen (University of California, Santa Barbara)

Taming Stable Diffusion for Text to 360 Panorama Image Generation 6347
Cheng Zhang (Monash University, Australia; Building 4.0 CRC, Australia), Qianyi Wu (Monash University, Australia), Camilo Cruz Gambardella (Monash University, Australia; Building 4.0 CRC, Australia), Xiaoshui Huang (Shanghai AI Laboratory, China), Dinh Phung (Monash University, Australia), Wanli Ouyang (Shanghai AI Laboratory, China), and Jianfei Cai (Monash University, Australia)

EmoGen: Emotional Image Content Generation with Text-to-Image Diffusion Models 6358
Jingyuan Yang (Shenzhen University, China), Jiawei Feng (Shenzhen University, China), and Hui Huang (Shenzhen University, China)

Carve3D: Improving Multi-view Reconstruction Consistency for Diffusion Models with RL Finetuning 6369
Desai Xie (Stony Brook University, USA), Jiahao Li (TTIC, USA), Hao Tan (Adobe Research, USA), Xin Sun (Adobe Research, USA), Zhixin Shu (Adobe Research, USA), Yi Zhou (Adobe Research, USA), Sai Bi (Adobe Research, USA), Sören Pirk (Kiel University, Germany), and Arie E. Kaufman (Stony Brook University, USA)

Move Anything with Layered Scene Diffusion 6380
Jiawei Ren (Nanyang Technological University), Mengmeng Xu (Meta AI), Jui-Chieh Wu (Meta AI), Ziwei Liu (Nanyang Technological University), Tao Xiang (Meta AI), and Antoine Toisoul (Meta AI)

DiffAgent: Fast and Accurate Text-to-Image API Selection with Large Language Model 6390
Lirui Zhao (Xiamen University, China; Shanghai AI Laboratory, China), Yue Yang (Shanghai Jiao Tong University, China; Shanghai AI Laboratory, China), Kaipeng Zhang (Shanghai AI Laboratory, China), Wenqi Shao (Shanghai AI Laboratory, China), Yuxin Zhang (Xiamen University, China), Yu Qiao (Shanghai AI Laboratory, China), Ping Luo (The University of Hong Kong, HKSAR; Shanghai AI Laboratory, China), and Rongrong Ji (Xiamen University, China)

CapHuman: Capture Your Moments in Parallel Universes 6400
Chao Liang (Zhejiang University), Fan Ma (Zhejiang University), Linchao Zhu (Zhejiang University), Yingying Deng (Huawei Technologies Ltd.), and Yi Yang (Zhejiang University)

IQ-VFI: Implicit Quadratic Motion Estimation for Video Frame Interpolation 6410
Mengshun Hu (Wuhan University), Kui Jiang (Harbin Institute of Technology), Zhihang Zhong (Shanghai Artificial Intelligence Laboratory), Zheng Wang (Wuhan University), and Yinqiang Zheng (The University of Tokyo)

Coarse-to-Fine Latent Diffusion for Pose-Guided Person Image Synthesis 6420
Yanzuo Lu (Sun Yat-sen University), Manlin Zhang (Sun Yat-sen University), Andy J Ma (Sun Yat-sen University), Xiaohua Xie (Sun Yat-sen University), and Jianhuang Lai (Sun Yat-sen University)

MACE: Mass Concept Erasure in Diffusion Models 6430
Shilin Lu (Nanyang Technological University, Singapore), Zilan Wang (Nanyang Technological University, Singapore), Leyang Li (Nanyang Technological University, Singapore), Yanzhu Liu (Agency for Science, Technology and Research, Singapore), and Adams Wai-Kin Kong (Nanyang Technological University, Singapore)

GenTron: Diffusion Transformers for Image and Video Generation 6441
Shoufa Chen (The University of Hong Kong), Mengmeng Xu (Meta), Jiawei Ren (Meta), Yuren Cong (Meta), Sen He (Meta), Yanping Xie (Meta), Animesh Sinha (Meta), Ping Luo (The University of Hong Kong), Tao Xiang (Meta), and Juan-Manuel Perez-Rua (Meta)

Relightful Harmonization: Lighting-aware Portrait Background Replacement 6452
Mengwei Ren (Adobe), Wei Xiong (Adobe), Jae Shin Yoon (Adobe), Zhixin Shu (Adobe), Jianming Zhang (Adobe), HyunJoon Jung (Adobe), Guido Gerig (New York University), and He Zhang (Adobe)

InstructVideo: Instructing Video Diffusion Models with Human Feedback 6463
Hangjie Yuan (Zhejiang University), Shiwei Zhang (Alibaba Group), Xiang Wang (Alibaba Group), Yujie Wei (Alibaba Group), Tao Feng (Tsinghua University), Yining Pan (Singapore University of Technology and Design), Yingya Zhang (Alibaba Group), Ziwei Liu (S-Lab, Nanyang Technological University), Samuel Albanie (CAML Lab, University of Cambridge), and Dong Ni (Zhejiang University)

SportsSloMo: A New Benchmark and Baselines for Human-centric Video Frame Interpolation .... N/A
Jiaben Chen (University of California, San Diego) and Huaizu Jiang (Northeastern University)

TeTriRF: Temporal Tri-Plane Radiance Fields for Efficient Free-Viewpoint Video 6487
Minye Wu (KU Leuven), Zehao Wang (KU Leuven), Georgios Kouros (KU Leuven), and Tinne Tuytelaars (KU Leuven)

SmartMask: Context Aware High-Fidelity Mask Generation for Fine-grained Object Insertion and Layout Control 6497
Jaskirat Singh (Australian National University, Australia), Jianming Zhang (Adobe Research, USA), Qing Liu (Adobe Research, USA), Cameron Smith (Adobe Research, USA), Zhe Lin (Adobe Research, USA), and Liang Zheng (Australian National University, Australia)

RAVE: Randomized Noise Shuffling for Fast and Consistent Video Editing with Diffusion Models 6507
Ozgur Kara (Georgia Institute of Technology, USA), Bariscan Kurtkaya (KUIS AI Center, Turkey), Hidir Yesiltepe (Virginia Polytechnic Institute and State University, USA), James M. Rehg (University of Illinois Urbana-Champaign, USA), and Pinar Yanardag (Virginia Polytechnic Institute and State University, USA)

LucidDreamer: Towards High-Fidelity Text-to-3D Generation via Interval Score Matching 6517
Yixun Liang (The Hong Kong University of Science and Technology (Guangzhou), China), Xin Yang (The Hong Kong University of Science and Technology, Hong Kong SAR, China; The Hong Kong University of Science and Technology (Guangzhou), China), Jiantao Lin (The Hong Kong University of Science and Technology (Guangzhou), China), Haodong Li (The Hong Kong University of Science and Technology (Guangzhou), China), Xiaogang Xu (Chinese University of Hong Kong, Hong Kong SAR, China; Zhejiang University, China), and Yingcong Chen (The Hong Kong University of Science and Technology, Hong Kong SAR, China; The Hong Kong University of Science and Technology (Guangzhou), China)

HyperDreamBooth: HyperNetworks for Fast Personalization of Text-to-Image Models 6527
Nataniel Ruiz (Google), Yuanzhen Li (Google), Varun Jampani (Google), Wei Wei (Google), Tingbo Hou (Google), Yael Pritch (Google), Neal Wadhwa (Google), Michael Rubinstein (Google), and Kfir Aberman (Google)

DreamVideo: Composing Your Dream Videos with Customized Subject and Motion 6537
Yujie Wei (Fudan University, China), Shiwei Zhang (Alibaba Group, China), Zhiwu Qing (Alibaba Group, China), Hangjie Yuan (Alibaba Group, China), Zhiheng Liu (Alibaba Group, China), Yu Liu (Alibaba Group, China), Yingya Zhang (Alibaba Group, China), Jingren Zhou (Alibaba Group, China), and Hongming Shan (Fudan University, China)

SurMo: Surface-based 4D Motion Modeling for Dynamic Human Rendering 6550
Tao Hu (Nanyang Technological University), Fangzhou Hong (Nanyang Technological University), and Ziwei Liu (Nanyang Technological University)

GenHowTo: Learning to Generate Actions and State Transformations from Instructional Videos.. 6561
Tomáš Souček (Czech Technical University), Dima Damen (University of Bristol), Michael Wray (University of Bristol), Ivan Laptev (Mohamed bin Zayed University of Artificial Intelligence), and Josef Sivic (Czech Technical University)

A Recipe for Scaling up Text-to-Video Generation with Text-free Videos 6572
Xiang Wang (Huazhong University of Science and Technology, China), Shiwei Zhang (Alibaba Group, China), Hangjie Yuan (Zhejiang University, China), Zhiwu Qing (Huazhong University of Science and Technology, China), Biao Gong (Alibaba Group, China), Yingya Zhang (Alibaba Group, China), Yujun Shen (Ant Group, China), Changxin Gao (Huazhong University of Science and Technology, China), and Nong Sang (Huazhong University of Science and Technology, China)

WaveFace: Authentic Face Restoration with Efficient Frequency Recovery 6583
Yunqi Miao (University of Warwick, UK), Jiankang Deng (Imperial College London), and Jungong Han (University of Sheffield)

AnyDoor: Zero-shot Object-level Image Customization 6593
Xi Chen (The University of Hong Kong), Lianghua Huang (Alibaba Group), Yu Liu (Alibaba Group), Yujun Shen (Ant Group), Deli Zhao (Alibaba Group), and Hengshuang Zhao (The University of Hong Kong)

ElasticDiffusion: Training-free Arbitrary Size Image Generation through Global-Local Content Separation 6603
Moayed Haji-Ali (Rice University), Guha Balakrishnan (Rice University), and Vicente Ordonez (Rice University)

One-step Diffusion with Distribution Matching Distillation 6613
Tianwei Yin (Massachusetts Institute of Technology, USA), Michaël Gharbi (Adobe Research, USA), Richard Zhang (Adobe Research, USA), Eli Shechtman (Adobe Research, USA), Frédo Durand (Massachusetts Institute of Technology, USA), William T. Freeman (Massachusetts Institute of Technology, USA), and Taesung Park (Adobe Research, USA)

Check, Locate, Rectify: A Training-Free Layout Calibration System for Text-to-Image Generation 6624
Biao Gong (Alibaba Group), Siteng Huang (Zhejiang University), Yutong Feng (Alibaba Group), Shiwei Zhang (Alibaba Group), Yuyuan Li (Zhejiang University), and Yu Liu (Alibaba Group)

Hierarchical Spatio-temporal Decoupling for Text-to-Video Generation 6635
Zhiwu Qing (Huazhong University of Science and Technology), Shiwei Zhang (Alibaba Group), Jiayu Wang (Alibaba Group), Xiang Wang (Huazhong University of Science and Technology), Yujie Wei (Fudan University), Yingya Zhang (Alibaba Group), Changxin Gao (Huazhong University of Science and Technology), and Nong Sang (Huazhong University of Science and Technology)

HumanGaussian: Text-Driven 3D Human Generation with Gaussian Splatting 6646
Xian Liu (The Chinese University of Hong Kong, Hong Kong), Xiaohang Zhan (Tencent AI Lab, China), Jiaxiang Tang (Peking University, China), Ying Shan (Tencent AI Lab, China), Gang Zeng (Peking University, China), Dahua Lin (The Chinese University of Hong Kong, Hong Kong), Xihui Liu (University of Hong Kong, Hong Kong), and Ziwei Liu (Nanyang Technological University, Singapore)

WonderJourney: Going from Anywhere to Everywhere 6658
Hong-Xing Yu (Stanford University), Haoyi Duan (Stanford University), Junhwa Hur (Google Research), Kyle Sargent (Stanford University), Michael Rubinstein (Google Research), William T. Freeman (Google Research), Forrester Cole (Google Research), Deqing Sun (Google Research), Noah Snavely (Google Research), Jiajun Wu (Stanford University), and Charles Herrmann (Google Research)

Balancing Act: Distribution-Guided Debiasing in Diffusion Models 6668
Rishubh Parihar (Indian Institute of Science, Bangalore), Abhijnya Bhat (Indian Institute of Science, Bangalore), Abhipsa Basu (Indian Institute of Science, Bangalore), Saswat Mallick (Indian Institute of Science, Bangalore), Jogendra Nath Kundu (Meta Reality Labs), and R. Venkatesh Babu (Indian Institute of Science, Bangalore)

SIGNeRF: Scene Integrated Generation for Neural Radiance Fields 6679
Jan-Niklas Dihlmann (University of Tübingen, Germany), Andreas Engelhardt (University of Tübingen, Germany), and Hendrik Lensch (University of Tübingen, Germany)

VideoBooth: Diffusion-based Video Generation with Image Prompts 6689
Yuming Jiang (S-Lab, Nanyang Technological University, Singapore), Tianxing Wu (S-Lab, Nanyang Technological University, Singapore), Shuai Yang (Wangxuan Institute of Computer Technology, Peking University, China), Chenyang Si (S-Lab, Nanyang Technological University, Singapore), Dahua Lin (Shanghai AI Laboratory, China), Yu Qiao (Shanghai AI Laboratory, China), Chen Change Loy (S-Lab, Nanyang Technological University, Singapore), and Ziwei Liu (S-Lab, Nanyang Technological University, Singapore)

Total Selfie: Generating Full-Body Selfies 6701
Bowei Chen (University of Washington, USA), Brian Curless (University of Washington, USA), Ira Kemelmacher-Shlizerman (University of Washington, USA), and Steven M. Seitz (University of Washington, USA)

CCEdit: Creative and Controllable Video Editing via Diffusion Models 6712
Ruoyu Feng (University of Science and Technology of China, China), Wenming Weng (University of Science and Technology of China, China), Yanhui Wang (University of Science and Technology of China, China), Yuhui Yuan (Microsoft Research Asia, China), Jianmin Bao (Microsoft Research Asia, China), Chong Luo (Microsoft Research Asia, China), Zhibo Chen (University of Science and Technology of China, China), and Baining Guo (Microsoft Research Asia, China)

Cinematic Behavior Transfer via NeRF-based Differentiable Filming 6723
Xuekun Jiang (Shanghai AI Lab, China), Anyi Rao (Stanford University, USA), Jingbo Wang (Shanghai AI Lab, China), Dahua Lin (Shanghai AI Lab, China; Chinese University of Hong Kong, China), and Bo Dai (Shanghai AI Lab, China)

Improving Subject-Driven Image Synthesis with Subject-Agnostic Guidance 6733
Kelvin C.K. Chan (Google), Yang Zhao (Google), Xuhui Jia (Google), Ming-Hsuan Yang (Google), and Huisheng Wang (Google)

Drag Your Noise: Interactive Point-based Editing via Diffusion Semantic Propagation 6743
Haofeng Liu (Singapore Management University, Singapore; South China Normal University, China), Chenshu Xu (Singapore Management University, Singapore), Yifei Yang (Singapore Management University, Singapore), Lihua Zeng (South China Normal University, China), and Shengfeng He (Singapore Management University, Singapore)

Learning Continuous 3D Words for Text-to-Image Generation 6753
Ta-Ying Cheng (University of Oxford, UK), Matheus Gadelha (Adobe Research, USA), Thibault Groueix (Adobe Research, USA), Matthew Fisher (Adobe Research, USA), Radomir Mech (Adobe Research, USA), Andrew Markham (University of Oxford, UK), and Niki Trigoni (University of Oxford, UK)

CHAIN: Enhancing Generalization in Data-Efficient GANs via lipsCHitz continuity constrAIned Normalization 6763
Yao Ni (The Australian National University) and Piotr Koniusz (Data61/CSIRO)

ViVid-1-to-3: Novel View Synthesis with Video Diffusion Models 6775
Jeong-gi Kwak (University of British Columbia, Canada; Korea University, South Korea), Erqun Dong (University of British Columbia, Canada; McGill University, Canada; Haiper Ltd., UK), Yuhe Jin (University of British Columbia, Canada), Hanseok Ko (Korea University, South Korea), Shweta Mahajan (University of British Columbia, Canada; Vector Institute for AI, Canada), and Kwang Moo Yi (University of British Columbia, Canada; Haiper Ltd., UK)

JeDi: Joint-Image Diffusion Models for Finetuning-Free Personalized Text-to-Image Generation 6786
Yu Zeng (Johns Hopkins University), Vishal M. Patel (Johns Hopkins University), Haochen Wang (TTI-Chicago), Xun Huang (NVIDIA Research), Ting-Chun Wang (NVIDIA Research), Ming-Yu Liu (NVIDIA Research), and Yogesh Balaji (NVIDIA Research)

GaussianDreamer: Fast Generation from Text to 3D Gaussians by Bridging 2D and 3D Diffusion Models 6796
Taoran Yi (Huazhong University of Science and Technology, China), Jiemin Fang (Huawei Inc., China), Junjie Wang (Huawei Inc., China), Guanjun Wu (Huazhong University of Science and Technology, China), Lingxi Xie (Huawei Inc., China), Xiaopeng Zhang (Huawei Inc., China), Wenyu Liu (Huazhong University of Science and Technology, China), Qi Tian (Huawei Inc., China), and Xinggang Wang (Huazhong University of Science and Technology, China)

Prompting Hard or Hardly Prompting: Prompt Inversion for Text-to-Image Diffusion Models 6808
Shweta Mahajan (University of British Columbia, Canada; Vector Institute for AI, Canada), Tanzila Rahman (University of British Columbia, Canada; Vector Institute for AI, Canada), Kwang Moo Yi (University of British Columbia, Canada), and Leonid Sigal (University of British Columbia, Canada; Vector Institute for AI, Canada)

MIGC: Multi-Instance Generation Controller for Text-to-Image Synthesis 6818
Dewei Zhou (Zhejiang University, China), You Li (Zhejiang University, China), Fan Ma (Zhejiang University, China), Xiaoting Zhang (Huawei Technologies Ltd., China), and Yi Yang (Zhejiang University, China)

Towards Text-guided 3D Scene Composition 6829
Qihang Zhang (The Chinese University of Hong Kong), Chaoyang Wang (Snap), Aliaksandr Siarohin (Snap), Peiye Zhuang (Snap), Yinghao Xu (Stanford University), Ceyuan Yang (The Chinese University of Hong Kong), Dahua Lin (The Chinese University of Hong Kong), Bolei Zhou (University of California, Los Angeles), Sergey Tulyakov (Snap), and Hsin-Ying Lee (Snap)

BerfScene: Bev-conditioned Equivariant Radiance Fields for Infinite 3D Scene Generation 6839
Qihang Zhang (The Chinese University of Hong Kong), Yinghao Xu (Stanford University), Yujun Shen (Ant Research), Bo Dai (Shanghai AI Lab), Bolei Zhou (University of California, Los Angeles), and Ceyuan Yang (Shanghai AI Lab)

Face2Diffusion for Fast and Editable Face Personalization 6850
Kaede Shiohara (The University of Tokyo) and Toshihiko Yamasaki (The University of Tokyo)

FreeDrag: Feature Dragging for Reliable Point-based Image Editing 6860
Pengyang Ling (University of Science and Technology of China), Lin Chen (University of Science and Technology of China), Pan Zhang (Shanghai AI Laboratory), Huaian Chen (University of Science and Technology of China), Yi Jin (University of Science and Technology of China), and Jinjin Zheng (University of Science and Technology of China)

OmniLocalRF: Omnidirectional Local Radiance Fields from Dynamic Videos 6871
Dongyoung Choi (KAIST), Hyeonjoong Jang (KAIST), and Min H. Kim (KAIST)

DIRECT-3D: Learning Direct Text-to-3D Generation on Massive Noisy 3D Data 6881
Qihao Liu (Johns Hopkins University, USA), Yi Zhang (Johns Hopkins University, USA), Song Bai (ByteDance, China), Adam Kortylewski (Max Planck Institute for Informatics, Germany), and Alan Yuille (Johns Hopkins University, USA)

Generate Like Experts: Multi-Stage Font Generation by Incorporating Font Transfer Process into Diffusion Models 6892
Bin Fu (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences), Fanghua Yu (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences), Anran Liu (The University of Hong Kong), Zixuan Wang (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences), Jie Wen (Harbin Institute of Technology, Shenzhen), Junjun He (Shanghai Artificial Intelligence Laboratory), and Yu Qiao (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences; Shanghai Artificial Intelligence Laboratory)

Panacea: Panoramic and Controllable Video Generation for Autonomous Driving 6902
Yuqing Wen (University of Science and Technology of China, China), Yucheng Zhao (Megvii Technology, China), Yingfei Liu (Megvii Technology, China), Fan Jia (Megvii Technology, China), Yanhui Wang (University of Science and Technology of China, China), Chong Luo (University of Science and Technology of China, China), Chi Zhang (Mach Drive, China), Tiancai Wang (Megvii Technology, China), Xiaoyan Sun (University of Science and Technology of China, China), and Xiangyu Zhang (Megvii Technology, China)

360DVD: Controllable Panorama Video Generation with 360-Degree Video Diffusion Model 6913
Qian Wang (Peking University), Weiqi Li (Peking University), Chong Mou (Peking University), Xinhua Cheng (Peking University), and Jian Zhang (Peking University)

CLiC: Concept Learning in Context 6924
Mehdi Safaee (Simon Fraser University), Aryan Mikaeili (Simon Fraser University), Or Patashnik (Tel Aviv University), Daniel Cohen-Or (Tel Aviv University), and Ali Mahdavi-Amiri (Simon Fraser University)

Z*: Zero-shot Style Transfer via Attention Reweighting 6934
Yingying Deng (Institute of Automation Chinese Academy of Sciences, China), Xiangyu He (Institute of Automation Chinese Academy of Sciences, China), Fan Tang (Institute of Computing Technology Chinese Academy of Sciences, China), and Weiming Dong (Institute of Automation Chinese Academy of Sciences, China)

Tackling the Singularities at the Endpoints of Time Intervals in Diffusion Models 6945
Pengze Zhang (Sun Yat-sen University), Hubery Yin (Wechat, Tencent Inc.), Chen Li (Wechat, Tencent Inc.), and Xiaohua Xie (Sun Yat-sen University)

CosmicMan: A Text-to-Image Foundation Model for Humans 6955
Shikai Li (Shanghai AI Laboratory, China), Jianglin Fu (Shanghai AI Laboratory, China), Kaiyuan Liu (Shanghai AI Laboratory, China), Wentao Wang (Shanghai AI Laboratory, China), Kwan-Yee Lin (Shanghai AI Laboratory, China), and Wayne Wu (Shanghai AI Laboratory, China)

Customize your NeRF: Adaptive Source Driven 3D Scene Editing via Local-Global Iterative Training 6966
Runze He (Institute of Information Engineering, Chinese Academy of Sciences, China; School of Cyber Security, University of Chinese Academy of Sciences, China), Shaofei Huang (Institute of Information Engineering, Chinese Academy of Sciences, China; School of Cyber Security, University of Chinese Academy of Sciences, China), Xuecheng Nie (Meitu Inc., China), Tianrui Hui (Hefei University of Technology, China), Luoqi Liu (Meitu Inc., China), Jiao Dai (Institute of Information Engineering, Chinese Academy of Sciences, China; School of Cyber Security, University of Chinese Academy of Sciences, China), Jizhong Han (Institute of Information Engineering, Chinese Academy of Sciences, China; School of Cyber Security, University of Chinese Academy of Sciences, China), Guanbin Li (Sun Yat-sen University, China), and Si Liu (Beihang University, China)

PICTURE: PhotorealistIC virtual Try-on from UnconstRained dEsigns 6976
Shuliang Ning (The Chinese University of HongKong, Shenzhen), Duomin Wang (Xiaobing.AI), Yipeng Qin (Cardiff University), Zirong Jin (The Chinese University of HongKong, Shenzhen), Baoyuan Wang (Xiaobing.AI), and Xiaoguang Han (The Chinese University of HongKong, Shenzhen)

Focus on Your Instruction: Fine-grained and Multi-instruction Image Editing by Attention Modulation 6986
Qin Guo (Peking University) and Tianwei Lin (Horizon Robotics)

Make-Your-Anchor: A Diffusion-based 2D Avatar Generation Framework 6997
Ziyao Huang (Chinese Academy of Sciences), Fan Tang (Chinese Academy of Sciences), Yong Zhang (Tencent AI Lab), Xiaodong Cun (Tencent AI Lab), Juan Cao (Chinese Academy of Sciences), Jintao Li (Chinese Academy of Sciences), and Tong-Yee Lee (National Cheng-Kung University)

Revisiting Non-Autoregressive Transformers for Efficient Image Synthesis 7007
Zanlin Ni (Tsinghua University, China), Yulin Wang (Tsinghua University, China), Renping Zhou (Tsinghua University, China), Jiayi Guo (Tsinghua University, China), Jinyi Hu (Tsinghua University, China), Zhiyuan Liu (Tsinghua University, China), Shiji Song (Tsinghua University, China), Yuan Yao (National University of Singapore, Singapore), and Gao Huang (Tsinghua University, China)

Texture-Preserving Diffusion Models for High-Fidelity Virtual Try-On 7017
Xu Yang (South China University of Technology, China), Changxing Ding (South China University of Technology, China), Zhibin Hong (S Research, USA), Junhao Huang (S Research, USA), Jin Tao (South China University of Technology, China), and Xiangmin Xu (South China University of Technology, China)

PromptCoT: Align Prompt Distribution via Adapted Chain-of-Thought 7027
Junyi Yao (Peking University), Yijiang Liu (Nanjing University), Zhen Dong (University of California, Berkeley), Mingfei Guo (Stanford), Helan Hu (Peking University), Kurt Keutzer (University of California, Berkeley), Li Du (Nanjing University), Daquan Zhou (Bytedance Inc. San Jose), and Shanghang Zhang (Peking University)

Snap Video: Scaled Spatiotemporal Transformers for Text-to-Video Synthesis 7038
Willi Menapace (Snap Inc., University of Trento), Aliaksandr Siarohin (Snap Inc.), Ivan Skorokhodov (Snap Inc.), Ekaterina Deyneka (Snap Inc.), Tsai-Shien Chen (UC Merced, Snap Inc.), Anil Kag (Snap Inc.), Yuwei Fang (Snap Inc.), Aleksei Stoliar (Snap Inc.), Elisa Ricci (University of Trento, Fondazione Bruno Kessler), Jian Ren (Snap Inc.), and Sergey Tulyakov (Snap Inc.)

L-MAGIC: Language Model Assisted Generation of Images with Coherence 7049
Zhipeng Cai (Intel Labs), Matthias Mueller (Intel Labs), Reiner Birkl (Intel Labs), Diana Wofk (Intel Labs), Shao-Yen Tseng (Intel Labs), Junda Cheng (Intel Labs), Gabriela Ben-Melech Stan (Intel Labs), Vasudev Lai (Intel Labs), and Michael Paulitsch (Intel Labs)

Text-Driven Image Editing via Learnable Regions 7059
Yuanze Lin (University of Oxford), Yi-Wen Chen (UC Merced), Yi-Hsuan Tsai (Google), Lu Jiang (Google), and Ming-Hsuan Yang (UC Merced)

On Exact Inversion of DPM-Solvers 7069
Seongmin Hong (Seoul National University, Republic of Korea), Kyeonghyun Lee (Seoul National University, Republic of Korea), Suh Yoon Jeon (Seoul National University, Republic of Korea), Hyewon Bae (Seoul National University, Republic of Korea), and Se Young Chun (Seoul National University, Republic of Korea)

ConsistNet: Enforcing 3D Consistency for Multi-view Images Diffusion 7079
Jiayu Yang (Australian National University, Australia), Ziang Cheng (Australian National University, Australia), Yunfei Duan (Tencent, Australia), Pan Ji (Tencent, China), and Hongdong Li (Australian National University, Australia)

LAMP: Learn A Motion Pattern for Few-Shot Video Generation 7089
Ruiqi Wu (VCIP, CS, Nankai University; MEGVII Technology), Liangyu Chen (MEGVII Technology), Tong Yang (MEGVII Technology), Chunle Guo (NKIARI, Shenzhen Futian; VCIP, CS, Nankai University), Chongyi Li (NKIARI, Shenzhen Futian; VCIP, CS, Nankai University), and Xiangyu Zhang (MEGVII Technology)

Task-Customized Mixture of Adapters for General Image Fusion 7099
Pengfei Zhu (Tianjin University, China), Yang Sun (Tianjin University, China), Bing Cao (Tianjin University, China), and Qinghua Hu (Tianjin University, China)

Beyond Textual Constraints: Learning Novel Diffusion Conditions with Fewer Examples 7109
Yuyang Yu (South China University of Technology, China), Bangzhen Liu (South China University of Technology, China), Chenxi Zheng (South China University of Technology, China), Xuemiao Xu (South China University of Technology, China), Huaidong Zhang (South China University of Technology, China), and Shengfeng He (Singapore Management University, Singapore)

Portrait4D: Learning One-Shot 4D Head Avatar Synthesis using Synthetic Data 7119
Yu Deng (Xiaobing.AI, China), Duomin Wang (Xiaobing.AI, China), Xiaohang Ren (Xiaobing.AI, China), Xingyu Chen (Xiaobing.AI, China), and Baoyuan Wang (Xiaobing.AI, China)

Animating General Image with Large Visual Motion Model 7131
Dengsheng Chen (Meituan, China), Xiaoming Wei (Meituan, China), and Xiaolin Wei (Meituan, China)

Sat2Scene: 3D Urban Scene Generation from Satellite Images with Diffusion 7141
Zuoyue Li (ETH Zürich), Zhenqiang Li (The University of Tokyo), Zhaopeng Cui (Zhejiang University), Marc Pollefeys (ETH Zürich, Microsoft), and Martin R. Oswald (University of Amsterdam)

Seeing and Hearing: Open-domain Visual-Audio Generation with Diffusion Latent Aligners 7151
Yazhou Xing (Hong Kong University of Science and Technology), Yingqing He (HKUST), Zeyue Tian (HKUST), Xintao Wang (Tencent), and Qifeng Chen (HKUST)

AVID: Any-Length Video Inpainting with Diffusion Model 7162
Zhixing Zhang (Rutgers University), Bichen Wu (GenAI, Meta), Xiaoyan Wang (MIT), Yaqiao Luo (GenAI, Meta), Luxin Zhang (GenAI, Meta), Yinan Zhao (GenAI, Meta), Peter Vajda (GenAI, Meta), Dimitris Metaxas (Rutgers University), and Licheng Yu (GenAI, Meta)

Generative Powers of Ten 7173
Xiaojuan Wang (University of Washington, USA), Janne Kontkanen (Google Research, USA), Brian Curless (University of Washington, Google Research, USA), Steven M. Seitz (University of Washington, Google Research, USA), Ira Kemelmacher-Shlizerman (University of Washington, Google Research, USA), Ben Mildenhall (Google Research, USA), Pratul Srinivasan (Google Research, USA), Dor Verbin (Google Research, USA), and Aleksander Holynski (UC Berkeley, Google Research, USA)

DistriFusion: Distributed Parallel Inference for High-Resolution Diffusion Models 7183
Muyang Li (MIT), Tianle Cai (Princeton), Jiaxin Cao (Lepton AI), Qinsheng Zhang (NVIDIA), Han Cai (MIT), Junjie Bai (Lepton AI), Yangqing Jia (Lepton AI), Kai Li (Princeton), and Song Han (MIT, NVIDIA)

Condition-Aware Neural Network for Controlled Image Generation 7194
Han Cai (MIT, USA), Muyang Li (MIT, USA), Qinsheng Zhang (NVIDIA, USA), Ming-Yu Liu (NVIDIA, USA), and Song Han (MIT, USA)

It's All About Your Sketch: Democratising Sketch Control in Diffusion Models 7204
Subhadeep Koley (University of Surrey, United Kingdom; iFlyTek-Surrey Joint Research Centre on Artificial Intelligence), Ayan Kumar Bhunia (University of Surrey, United Kingdom), Deeptanshu Sekhri (University of Surrey, United Kingdom), Aneeshan Sain (University of Surrey, United Kingdom), Pinaki Nath Chowdhury (University of Surrey, United Kingdom), Tao Xiang (University of Surrey, United Kingdom; iFlyTek-Surrey Joint Research Centre on Artificial Intelligence), and Yi-Zhe Song (University of Surrey, United Kingdom; iFlyTek-Surrey Joint Research Centre on Artificial Intelligence)

FaceChain-SuDe: Building Derived Class to Inherit Category Attributes for One-shot Subject-Driven Generation 7215
Pengchong Qiao (Peking University, China), Lei Shang (Alibaba Group, China), Chang Liu (Tsinghua University, China), Baigui Sun (Alibaba Group, China), Xiangyang Ji (Tsinghua University, China), and Jie Chen (Peking University, China and Peng Cheng Laboratory, China)

In-N-Out: Faithful 3D GAN Inversion with Volumetric Decomposition for Face Editing 7225
Yiran Xu (University of Maryland, College Park, USA), Zhixin Shu (Adobe Research, USA), Cameron Smith (Adobe Research), Seoung Wug Oh (Adobe Research), and Jia-Bin Huang (University of Maryland, College Park, USA)

Video Prediction by Modeling Videos as Continuous Multi-Dimensional Processes 7236
Gaurav Shrivastava (University of Maryland College Park, USA) and Abhinav Shrivastava (University of Maryland College Park, USA)

DetDiffusion: Synergizing Generative and Perceptive Models for Enhanced Data Generation and Perception 7246
Yibo Wang (Tsinghua University), Ruiyuan Gao (The Chinese University of Hong Kong), Kai Chen (The Hong Kong University of Science and Technology), Kaiqiang Zhou (Huawei Technologies Ltd.), Yingjie Cai (Huawei Technologies Ltd.), Lanqing Hong (Huawei Technologies Ltd.), Zhenguo Li (Huawei Technologies Ltd.), Lihui Jiang (Huawei Technologies Ltd.), Dit-Yan Yeung (The Hong Kong University of Science and Technology), Qiang Xu (The Chinese University of Hong Kong), and Kai Zhang (Tsinghua University)

Structure-Guided Adversarial Training of Diffusion Models 7256
Ling Yang (Peking University), Haotian Qian (Peking University), Zhilong Zhang (Peking University), Jingwei Liu (Peking University), and Bin Cui (Peking University)

Learning Adaptive Spatial Coherent Correlations for Speech-Preserving Facial Expression Manipulation 7267
Tianshui Chen (Guangdong University of Technology), Jianman Lin (Guangdong University of Technology), Zhijing Yang (Guangdong University of Technology), Chunmei Qing (South China University of Technology), and Liang Lin (Sun Yat-Sen University)

On the Content Bias in Fréchet Video Distance 7277
Songwei Ge (University of Maryland, College Park), Aniruddha Mahapatra (Adobe Research), Gaurav Parmar (Carnegie Mellon University), Jun-Yan Zhu (Carnegie Mellon University), and Jia-Bin Huang (University of Maryland, College Park)

Residual Learning in Diffusion Models 7289
Junyu Zhang (Central South University), Daochang Liu (The University of Sydney), Eunbyung Park (Sungkyunkwan University), Shichao Zhang (Guangxi Normal University), and Chang Xu (The University of Sydney)

A Unified Approach for Text- and Image-guided 4D Scene Generation 7300
Yufeng Zheng (ETH Zurich, Switzerland; NVIDIA, Switzerland), Xueting Li (NVIDIA, USA), Koki Nagano (NVIDIA, USA), Sifei Liu (NVIDIA, USA), Otmar Hilliges (ETH Zurich, Switzerland), and Shalini De Mello (NVIDIA, USA)

VideoCrafter2: Overcoming Data Limitations for High-Quality Video Diffusion Models 7310
Haoxin Chen (Tencent AI Lab, China), Yong Zhang (Tencent AI Lab, China), Xiaodong Cun (Tencent AI Lab, China), Menghan Xia (Tencent AI Lab, China), Xintao Wang (Tencent AI Lab, China), Chao Weng (Tencent AI Lab, China), and Ying Shan (Tencent AI Lab, China)

Neural Implicit Morphing of Face Images 7321
Guilherme Schardong (Institute of Systems and Robotics, Portugal), Tiago Novello (Institute of Pure and Applied Mathematics, Brazil), Hallison Paz (Institute of Pure and Applied Mathematics, Brazil), Iurii Medvedev (Institute of Systems and Robotics, Portugal), Vinícius da Silva (Pontifical Catholic University of Rio de Janeiro, Brazil), Luiz Velho (Institute of Pure and Applied Mathematics, Brazil), and Nuno Gonçalves (University of Coimbra, Portugal)

One More Step: A Versatile Plug-and-Play Module for Rectifying Diffusion Schedule Flaws and Enhancing Low-Frequency Controls 7331
Minghui Hu (Nanyang Technological University, Singapore), Jianbin Zheng (South China University of Technology, China), Chuanxia Zheng (University of Oxford, UK), Chaoyue Wang (The University of Sydney, Australia), Dacheng Tao (Nanyang Technological University, Singapore), and Tat-Jen Cham (Nanyang Technological University, Singapore)

Video Interpolation with Diffusion Models 7341
Siddhant Jain (Google Research, USA), Daniel Watson (Google Deepmind, Canada), Eric Tabellion (Google Research, USA), Aleksander Hołynski (Google Research, USA), Ben Poole (Google Deepmind, USA), and Janne Kontkanen (Google Research, USA)

DiffSHEG: A Diffusion-Based Approach for Real-Time Speech-driven Holistic 3D Expression and Gesture Generation 7352
Junming Chen (Hong Kong University of Science and Technology), Yunfei Liu (International Digital Economy Academy), Jianan Wang (International Digital Economy Academy), Ailing Zeng (International Digital Economy Academy), Yu Li (International Digital Economy Academy), and Qifeng Chen (Hong Kong University of Science and Technology)

TFMQ-DM: Temporal Feature Maintenance Quantization for Diffusion Models 7362
Yushi Huang (Beihang University,SenseTime Research), Ruihao Gong (Beihang University, SenseTime Research), Jing Liu (Monash University, SenseTime Research), Tianlong Chen (UT Austin), and Xianglong Liu (Beihang University)

Improving Training Efficiency of Diffusion Models via Multi-Stage Framework and Tailored Multi-Decoder Architecture 7372
Huijie Zhang (University of Michigan, USA), Yifu Lu (University of Michigan, USA), Ismail Alkhouri (Michigan State University, USA; University of Michigan, USA), Saiprasad Ravishankar (Michigan State University, USA), Dogyoon Song (University of Michigan, USA), and Qing Qu (University of Michigan, USA)

Scaling Laws of Synthetic Images for Model Training ... for Now 7382
Lijie Fan (MIT CSAIL), Kaifeng Chen (Google Research), Dilip Krishnan (Google Research), Dina Katabi (MIT CSAIL), Phillip Isola (MIT CSAIL), and Yonglong Tian (Google Research)

BIVDiff: A Training-Free Framework for General-Purpose Video Synthesis via Bridging Image and Video Diffusion Models 7393
Fengyuan Shi (Nanjing University, China), Jiaxi Gu (Huawei Noah's Ark Lab, China), Hang Xu (Huawei Noah's Ark Lab, China), Songcen Xu (Huawei Noah's Ark Lab, China), Wei Zhang (Huawei Noah's Ark Lab, China), and Limin Wang (Nanjing University, China; Shanghai AI Laboratory, China)

MaskINT: Video Editing via Interpolative Non-autoregressive Masked Transformers 7403
Haoyu Ma (University of California, Irvine), Shahin Mahdizadehaghdam (GenAI, Meta), Bichen Wu (GenAI, Meta), Zhipeng Fan (GenAI, Meta), Yuchao Gu (National University of Singapore), Wenliang Zhao (GenAI, Meta), Lior Shapira (GenAI, Meta), and Xiaohui Xie (University of California, Irvine)

Pose Adapted Shape Learning for Large-Pose Face Reenactment 7413
Gee-Sern Jison Hsu (National Taiwan University of Science and Technology, Taipei, Taiwan), Jie-Ying Zhang (National Taiwan University of Science and Technology, Taipei, Taiwan), Huang Yu Hsiang (National Taiwan University of Science and Technology, Taipei, Taiwan), and Wei-Jie Hong (National Taiwan University of Science and Technology, Taipei, Taiwan)

PRDP: Proximal Reward Difference Prediction for Large-Scale Reward Finetuning of Diffusion Models 7423
Fei Deng (Google, USA; Rutgers University, USA), Qifei Wang (Google, USA), Wei Wei (Accenture, USA), Tingbo Hou (Google, USA), and Matthias Grundmann (Google, USA)

Discriminative Probing and Tuning for Text-to-Image Generation 7434
Leigang Qu (National University of Singapore), Wenjie Wang (National University of Singapore), Yongqi Li (Hong Kong Polytechnic University), Hanwang Zhang (Nanyang Technological University), Liqiang Nie (Harbin Institute of Technology (Shenzhen)), and Tat-Seng Chua (National University of Singapore)

Towards Automated Movie Trailer Generation 7445
Dawit Mureja Argaw (KAIST), Mattia Soldan (KAUST), Alejandro Pardo (KAUST), Chen Zhao (KAUST), Fabian Caba Heilbron (Adobe Research), Joon Son Chung (KAIST), and Bernard Ghanem (KAUST)

CDFormer: When Degradation Prediction Embraces Diffusion Model for Blind Image Super-Resolution 7455
Qingguo Liu (Nanjing University of Aeronautics And Astronautics, China), Chenyi Zhuang (Nanjing University of Aeronautics And Astronautics, China), Pan Gao (Nanjing University of Aeronautics And Astronautics, China), and Jie Qin (Nanjing University of Aeronautics And Astronautics, China)

FreeControl: Training-Free Spatial Control of Any Text-to-Image Diffusion Model with Any Condition 7465
Sicheng Mo (University of California, Los Angeles), Fangzhou Mu (University of Wisconsin-Madison), Kuan Heng Lin (University of California, Los Angeles), Yanli Liu (Innopeak Technology, Inc), Bochen Guan (Innopeak Technology, Inc), Yin Li (University of Wisconsin-Madison), and Bolei Zhou (University of California, Los Angeles)

RealCustom: Narrowing Real Text Word for Real-Time Open-Domain Text-to-Image Customization… 7476
Mengqi Huang (University of Science and Technology of China, China), Zhendong Mao (University of Science and Technology of China, China), Mingcong Liu (ByteDance Inc., China), Qian He (ByteDance Inc., China), and Yongdong Zhang (University of Science and Technology of China, China)

VidToMe: Video Token Merging for Zero-Shot Video Editing 7486
Xirui Li (Shanghai Jiao Tong University), Chao Ma (Shanghai Jiao Tong University), Xiaokang Yang (Shanghai Jiao Tong University), and Ming-Hsuan Yang (UC Merced)

Layout-Agnostic Scene Text Image Synthesis with Diffusion Models 7496
Qilong Zhangli (Rutgers University, USA), Jindong Jiang (Rutgers University, USA), Di Liu (Rutgers University, USA), Licheng Yu (Meta AI, USA), Xiaoliang Dai (Meta AI, USA), Ankit Ramchandani (Meta AI, USA), Guan Pang (Meta AI, USA), Dimitris N. Metaxas (Rutgers University, USA), and Praveen Krishnan (Meta AI, USA)

3D Multi-frame Fusion for Video Stabilization 7507
Zhan Peng (Huazhong University of Science and Technology, China), Xinyi Ye (Huazhong University of Science and Technology, China), Weiyue Zhao (Huazhong University of Science and Technology, China), Tianqi Liu (Huazhong University of Science and Technology, China), Huiqiang Sun (Huazhong University of Science and Technology, China), Baopu Li (Huazhong University of Science and Technology, China), and Zhiguo Cao (Huazhong University of Science and Technology, China)

DyBluRF: Dynamic Neural Radiance Fields from Blurry Monocular Video 7517
Huiqiang Sun (Huazhong University of Science and Technology, China), Xingyi Li (Huazhong University of Science and Technology, China), Liao Shen (Huazhong University of Science and Technology, China), Xinyi Ye (Huazhong University of Science and Technology, China), Ke Xian (Huazhong University of Science and Technology, China), and Zhiguo Cao (Huazhong University of Science and Technology, China)

A Video is Worth 256 Bases: Spatial-Temporal Expectation-Maximization Inversion for Zero-Shot Video Editing 7528
Maomao Li (The University of Hong Kong, China), Yu Li (International Digital Economy Academy (IDEA), China), Tianyu Yang (International Digital Economy Academy (IDEA), China), Yunfei Liu (International Digital Economy Academy (IDEA), China), Dongxu Yue (Peking University, China), Zhihui Lin (Tsinghua University, China), and Dong Xu (The University of Hong Kong, China)

StrokeFaceNeRF: Stroke-based Facial Appearance Editing in Neural Radiance Field 7538
Xiao-Juan Li (Institute of Computing Technology, Chinese Academy of Sciences; University of the Chinese Academy of Sciences), Dingxi Zhang (University of Chinese Academy of Sciences; Institute of Computing Technology, Chinese Academy of Sciences), Shu-Yu Chen (Institute of Computing Technology, Chinese Academy of Sciences), and Feng-Lin Liu (Institute of Computing Technology, Chinese Academy of Sciences; University of Chinese Academy of Sciences)

Smooth Diffusion: Crafting Smooth Latent Spaces in Diffusion Models 7548
Jiayi Guo (Georgia Tech, USA; Tsinghua University, China), Xingqian Xu (UIUC, USA; Picsart AI Research, USA), Yifan Pu (Tsinghua University, China), Zanlin Ni (Tsinghua University, China), Chaofei Wang (Tsinghua University, China), Manushree Vasu (Georgia Tech, USA), Shiji Song (Tsinghua University, China), Gao Huang (Tsinghua University, China), and Humphrey Shi (Georgia Tech & UIUC, USA; Picsart AI Research, USA)

One-dimensional Adapter to Rule Them All: Concepts, Diffusion Models and Erasing Applications 7559
Mengyao Lyu (Tsinghua University, BNRist), Yuhong Yang (Tsinghua University, BNRist), Haiwen Hong (Alibaba Group), Hui Chen (Tsinghua University, BNRist), Xuan Jin (Alibaba Group), Yuan He (Alibaba Group), Hui Xue (Alibaba Group), Jungong Han (Tsinghua University, BNRist), and Guiguang Ding (Tsinghua University, BNRist)

Hierarchical Patch Diffusion Models for High-Resolution Video Generation 7569
Ivan Skorokhodov (Snap Inc., USA), Willi Menapace (Snap Inc., USA), Aliaksandr Siarohin (Snap Inc., USA), and Sergey Tulyakov (Snap Inc., USA)

Taming the Tail in Class-Conditional GANs: Knowledge Sharing via Unconditional Training at Lower Resolutions 7580
Saeed Khorram (Oregon State University, USA), Mingqi Jiang (Oregon State University, USA), Mohamad Shahbazi (ETH Zurich, Switzerland), Mohamad H. Danesh (McGill University, Canada), and Li Fuxin (Oregon State University, USA)

Don't Look into the Dark: Latent Codes for Pluralistic Image Inpainting 7591
Haiwei Chen (University of Southern California) and Yajie Zhao (University of Southern California)

Content-Style Decoupling for Unsupervised Makeup Transfer without Generating Pseudo Ground Truth 7601
Zhaoyang Sun (Wuhan University of Technology, China), Shengwu Xiong (Wuhan University of Technology, China), Yaxiong Chen (Wuhan University of Technology, China), and Yi Rong (Wuhan University of Technology, China)

Generative Rendering: Controllable 4D-Guided Video Generation with 2D Diffusion Models 7611
Shengqu Cai (Stanford University), Duygu Ceylan (Adobe Research), Matheus Gadelha (Adobe Research), Chun-Hao Paul Huang (Adobe Research), Tuanfeng Yang Wang (Adobe Research), and Gordon Wetzstein (Stanford University)

VideoSwap: Customized Video Subject Swapping with Interactive Semantic Point Correspondence 7621
Yuchao Gu (National University of Singapore, Singapore), Yipin Zhou (Meta, USA), Bichen Wu (Meta, USA), Licheng Yu (Meta, USA), Jia-Wei Liu (National University of Singapore, Singapore), Rui Zhao (National University of Singapore, Singapore), Jay Zhangjie Wu (National University of Singapore, Singapore), David Junhao Zhang (National University of Singapore, Singapore), Mike Zheng Shou (National University of Singapore, Singapore), and Kevin Tang (Meta, USA)

Rethinking the Objectives of Vector-Quantized Tokenizers for Image Synthesis 7631
Yuchao Gu (National University of Singapore, Singapore), Xintao Wang (Tencent, China), Yixiao Ge (Tencent, China), Ying Shan (Tencent, China), and Mike Zheng Shou (National University of Singapore, Singapore)

Dysen-VDM: Empowering Dynamics-aware Text-to-Video Diffusion with LLMs 7641
Hao Fei (National University of Singapore, Singapore), Shengqiong Wu (National University of Singapore, Singapore), Wei Ji (National University of Singapore, Singapore), Hanwang Zhang (Nanyang Technological University, Singapore), and Tat-Seng Chua (National University of Singapore, Singapore)

Geometry-aware Reconstruction and Fusion-refined Rendering for Generalizable Neural Radiance Fields 7654
Tianqi Liu (Huazhong University of Science and Technology), Xinyi Ye (Huazhong University of Science and Technology), Min Shi (Huazhong University of Science and Technology), Zihao Huang (Huazhong University of Science and Technology), Zhiyu Pan (Huazhong University of Science and Technology), Zhan Peng (Huazhong University of Science and Technology), and Zhiguo Cao (Huazhong University of Science and Technology)

DynVideo-E: Harnessing Dynamic NeRF for Large-Scale Motion- and View-Change Human-Centric Video Editing 7664
Jia-Wei Liu (National University of Singapore, Singapore), Yan-Pei Cao (Tencent, China), Jay Zhangjie Wu (National University of Singapore, Singapore), Weijia Mao (National University of Singapore, Singapore), Yuchao Gu (National University of Singapore, Singapore), Rui Zhao (National University of Singapore, Singapore), Jussi Keppo (National University of Singapore, Singapore), Ying Shan (Tencent, China), and Mike Zheng Shou (National University of Singapore, Singapore)

High-fidelity Person-centric Subject-to-Image Synthesis 7675
Yibin Wang (Fudan University, China; Zhejiang University of Technology, China), Weizhong Zhang (Fudan University, China;), Jianwei Zheng (Zhejiang University of Technology, China), and Cheng Jin (Fudan University, China; Haina Lab, China)

Relation Rectification in Diffusion Model 7685
Yinwei Wu (National University of Singapore), Xingyi Yang (National University of Singapore), and Xinchao Wang (National University of Singapore)

Diffusion Handles Enabling 3D Edits for Diffusion Models by Lifting Activations to 3D 7695
Karran Pandey (University of Toronto), Paul Guerrero (Adobe Research), Matheus Gadelha (Adobe Research), Yannick Hold-Geoffroy (Adobe Research), Karan Singh (University of Toronto), and Niloy J. Mitra (Adobe Research, University College London)

LeftRefill: Filling Right Canvas based on Left Reference through Generalized Text-to-Image Diffusion Model 7705
Chenjie Cao (Fudan University), Yunuo Cai (Fudan University), Qiaole Dong (Fudan University), Yikai Wang (Fudan University), and Yanwei Fu (Fudan University)

FSRT: Facial Scene Representation Transformer for Face Reenactment from Factorized Appearance, Head-pose, and Facial Expression Features 7716
Andre Rochow (University of Bonn, Germany), Max Schwarz (University of Bonn, Germany), and Sven Behnke (University of Bonn, Germany)

Tailored Visions: Enhancing Text-to-Image Generation with Personalized Prompt Rewriting 7727
Zijie Chen (Westlake University), Lichao Zhang (Westlake University), Fangsheng Weng (Scietrain), Lili Pan (University of Electronic Science and Technology of China), and Zhenzhong Lan (Westlake University)

MMA-Diffusion: MultiModal Attack on Diffusion Models 7737
Yijun Yang (The Chinese University of Hong Kong, Hong Kong, S.A.R, China), Ruiyuan Gao (The Chinese University of Hong Kong, Hong Kong, S.A.R, China), Xiaosen Wang (Huawei Singular Security Lab, China), Tsung-Yi Ho (The Chinese University of Hong Kong, Hong Kong, S.A.R, China), Nan Xu (Institute of Automation Chinese Academy of Sciences, China), and Qiang Xu (The Chinese University of Hong Kong, Hong Kong, S.A.R, China)

PIA: Your Personalized Image Animator via Plug-and-Play Modules in Text-to-Image Models 7747
Yiming Zhang (Shanghai Artificial Intelligence Laboratory, Dalian University of Technology), Zhening Xing (Shanghai Artificial Intelligence Laboratory), Yanhong Zeng (Shanghai Artificial Intelligence Laboratory), Youqing Fang (Shanghai Artificial Intelligence Laboratory), and Kai Chen (Shanghai Artificial Intelligence Laboratory)

Codebook Transfer with Part-of-Speech for Vector-Quantized Image Modeling 7757
Baoquan Zhang (Harbin Institute of Technology, Shenzhen), Huaibin Wang (Harbin Institute of Technology, Shenzhen), Chuyao Luo (Harbin Institute of Technology, Shenzhen), Xutao Li (Harbin Institute of Technology, Shenzhen), Guotao Liang (Harbin Institute of Technology, Shenzhen), Yunming Ye (Harbin Institute of Technology, Shenzhen), Xiaochen Qi (ShenZhen SiFar Co., Ltd.), and Yao He (ShenZhen SiFar Co., Ltd.)

Generating Non-Stationary Textures using Self-Rectification 7767
Yang Zhou (Shenzhen University, China), Rongjun Xiao (Shenzhen University, China), Dani Lischinski (The Hebrew University of Jerusalem, Israel), Daniel Cohen-Or (Shenzhen University, China; Tel Aviv University, Israel), and Hui Huang (Shenzhen University, China)

Fast ODE-based Sampling for Diffusion Models in Around 5 Steps 7777
Zhenyu Zhou (Zhejiang University, China), Defang Chen (Zhejiang University, China), Can Wang (Zhejiang University, China), and Chun Chen (Zhejiang University, China)

Deformable One-shot Face Stylization via DINO Semantic Guidance 7787
Yang Zhou (Shenzhen University, China), Zichong Chen (Shenzhen University, China), and Hui Huang (Shenzhen University, China)

Learning Disentangled Identifiers for Action-Customized Text-to-Image Generation 7797
Siteng Huang (Zhejiang University; Westlake University), Biao Gong (Alibaba Group), Yutong Feng (Alibaba Group), Xi Chen (Alibaba Group), Yuqian Fu (ETH Zürich), Yu Liu (Alibaba Group), and Donglin Wang (Westlake Univerisity)

SwiftBrush: One-Step Text-to-Image Diffusion Model with Variational Score Distillation 7807
Thuan Hoang Nguyen (VinAI Research) and Anh Tran (VinAI Research)

Towards Understanding Cross and Self-Attention in Stable Diffusion for Text-Guided Image Editing 7817
Bingyan Liu (South China University of Technology), Chengyu Wang (Alibaba Group), Tingfeng Cao (South China University of Technology), Kui Jia (The Chinese University of Hong Kong, Shenzhen), and Jun Huang (Alibaba Group)

SimDA: Simple Diffusion Adapter for Efficient Video Generation 7827
Zhen Xing (Fudan University, China), Qi Dai (MicroSoft Research Asia), Han Hu (Microsoft Research Asia), Zuxuan Wu (Fudan University), and Yu-Gang Jiang (Fudan University)

Unlocking Pre-trained Image Backbones for Semantic Image Synthesis 7840
Tariq Berrada Ifriqi (Meta & INRIA, France), Jakob Verbeek (Meta, France), Camille Couprie (Meta, France), and Karteek Alahari (INRIA, France)

Shadow-Enlightened Image Outpainting 7850
Hang Yu (Shanghai University, China), Ruilin Li (Shanghai University, China), Shaorong Xie (Shanghai University, China), and Jiayan Qiu (Univerisity of Leicester, United Kingdom)

Exploiting Diffusion Prior for Generalizable Dense Prediction 7861
Hsin-Ying Lee (University of California, Merced), Hung-Yu Tseng (Meta), Hsin-Ying Lee (Snap Research), and Ming-Hsuan Yang (University of California, Merced)

StyleCineGAN: Landscape Cinemagraph Generation using a Pre-trained StyleGAN 7872
Jongwoo Choi (KAIST), Kwanggyoon Seo (KAIST), Amirsaman Ashtari (KAIST), and Junyong Noh (KAIST)

MotionEditor: Editing Video Motion via Content-Aware Diffusion 7882
Shuyuan Tu (Fudan University, China), Qi Dai (Microsoft Research Asia, China), Zhi-Qi Cheng (Carnegie Mellon University, America), Han Hu (Microsoft Research Asia, China), Xintong Han (Huya Inc, China), Zuxuan Wu (Fudan University, China), and Yu-Gang Jiang (Fudan University, China)

DanceCamera3D: 3D Camera Movement Synthesis with Music and Dance 7892
Zixuan Wang (Tsinghua University, China), Jia Jia (Tsinghua University, China), Shikun Sun (Tsinghua University, China), Haozhe Wu (Tsinghua University, China), Rong Han (Tsinghua University, China), Zhenyu Li (Tsinghua University, China), Di Tang (ByteDance Hangzhou, China), Jiaqing Zhou (ByteDance Hangzhou, China), and Jiebo Luo (University of Rochester, USA)

Diversity-aware Channel Pruning for StyleGAN Compression 7902
Jiwoo Chung (Sungkyunkwan University), Sangeek Hyun (Sungkyunkwan University), Sang-Heon Shim (Sungkyunkwan University), and Jae-Pil Heo (Sungkyunkwan University)

DiffMorpher: Unleashing the Capability of Diffusion Models for Image Morphing 7912
Kaiwen Zhang (Tsinghua University, China; Shangihai AI Laboratory, China), Yifan Zhou (Nanyang Technological University, Singapore), Xudong Xu (Shangihai AI Laboratory, China), Bo Dai (Shangihai AI Laboratory, China), and Xingang Pan (Nanyang Technological University, Singapore)

StegoGAN: Leveraging Steganography for Non-Bijective Image-to-Image Translation 7922
Sidi Wu (ETH Zurich), Yizi Chen (ETH Zurich), Samuel Mermet (Univ Gustave Eiffel), Lorenz Hurni (ETH Zurich), Konrad Schindler (ETH Zurich), Nicolas Gonthier (Univ Gustave Eiffel; IGN), and Loic Landrieu (Univ Gustave Eiffel)

Grounded Text-to-Image Synthesis with Attention Refocusing 7932
Quynh Phung (University of Maryland, College Park), Songwei Ge (University of Maryland, College Park), and Jia-Bin Huang (University of Maryland, College Park)

VecFusion: Vector Font Generation with Diffusion 7943
Vikas Thamizharasan (University of Massachusetts Amherst, USA), Difan Liu (Adobe Research, USA), Shantanu Agarwal (University of Massachusetts Amherst, USA), Matthew Fisher (Adobe Research, USA), Michael Gharbi (Adobe Research, USA), Oliver Wang (Google Research, USA), Alec Jacobson (Adobe Research, USA), and Evangelos Kalogerakis (University of Massachusetts Amherst, USA)

Single Mesh Diffusion Models with Field Latents for Texture Generation 7953
Thomas W. Mitchel (PlayStation, USA), Carlos Esteves (Google Research, USA), and Ameesh Makadia (Google Research, USA)

Orthogonal Adaptation for Modular Customization of Diffusion Models 7964
Ryan Po (Stanford University), Guandao Yang (Stanford University), Kfir Aberman (Snap Research), and Gordon Wetzstein (Stanford University)

Low-Latency Neural Stereo Streaming 7974
Qiqi Hou (Qualcomm Technologies, Inc), Farzad Farhadzadeh (Qualcomm Technologies, Inc), Amir Said (Qualcomm Technologies, Inc), Guillaume Sautiere (Qualcomm Technologies, Inc), and Hoang Le (Qualcomm Technologies, Inc)

TextCraftor: Your Text Encoder Can be Image Quality Controller 7985
Yanyu Li (Northeastern University, USA), Xian Liu (Snap Inc., USA), Anil Kag (Snap Inc., USA), Ju Hu (Snap Inc., USA), Yerlan Idelbayev (Snap Inc., USA), Dhritiman Sagar (Snap Inc., USA), Yanzhi Wang (Northeastern University, USA), Sergey Tulyakov (Snap Inc., USA), and Jian Ren (Snap Inc., USA)

4D-fy: Text-to-4D Generation Using Hybrid Score Distillation Sampling 7996
Sherwin Bahmani (University of Toronto, Canada ; Vector Institute, Canada), Ivan Skorokhodov (KAUST, Saudi Arabia ; Snap Inc., UAE), Victor Rong (University of Toronto, Canada ; Vector Institute, Canada), Gordon Wetzstein (Stanford University, USA), Leonidas Guibas (Stanford University, USA), Peter Wonka (KAUST, Saudi Arabia), Sergey Tulyakov (Snap Inc., USA), Jeong Joon Park (University of Michigan, USA), Andrea Tagliasacchi (University of Toronto, Canada ; SFU, Canada ; Google, Canada), and David B. Lindell (University of Toronto, Canada ; Vector Institute, Canada)

Image Neural Field Diffusion Models 8007
Yinbo Chen (UC San Diego), Oliver Wang (Google Research), Richard Zhang (Adobe Research), Eli Shechtman (Adobe Research), Xiaolong Wang (UC San Diego), and Michael Gharbi (Adobe Research)

Learning Multi-Dimensional Human Preference for Text-to-Image Generation 8018
Sixian Zhang (Kuaishou Technology, China), Bohan Wang (Kuaishou Technology, China), Junqiang Wu (Kuaishou Technology, China), Yan Li (Kuaishou Technology, China), Tingting Gao (Kuaishou Technology, China), Di Zhang (Kuaishou Technology, China), and Zhongyuan Wang (Kuaishou Technology, China)

Dynamic Policy-Driven Adaptive Multi-Instance Learning for Whole Slide Image Classification 8028
Tingting Zheng (Harbin Institute of Technology, China), Kui Jiang (Harbin Institute of Technology, China), and Hongxun Yao (Harbin Institute of Technology, China)

Structure Matters: Tackling the Semantic Discrepancy in Diffusion Models for Image Inpainting 8038
Haipeng Liu (Hefei University of Technology, China), Yang Wang (Hefei University of Technology, China), Biao Qian (Hefei University of Technology, China), Meng Wang (Hefei University of Technology, China), and Yong Rui (Lenovo Research, China)

IMPRINT: Generative Object Compositing by Learning Identity-Preserving Representation 8048
Yizhi Song (Purdue University, USA), Zhifei Zhang (Adobe, USA), Zhe Lin (Adobe, USA), Scott Cohen (Adobe, USA), Brian Price (Adobe, USA), Jianming Zhang (Adobe, USA), Soo Ye Kim (Adobe, USA), He Zhang (Adobe, USA), Wei Xiong (Adobe, USA), and Daniel Aliaga (Purdue University, USA)

Puff-Net: Efficient Style Transfer with Pure Content and Style Feature Fusion Network 8059
Sizhe Zheng (Nanjing University of Aeronautics and Astronautics, China), Pan Gao (Nanjing University of Aeronautics and Astronautics, China), Peng Zhou (Nanjing University of Aeronautics and Astronautics, China), and Jie Qin (Nanjing University of Aeronautics and Astronautics, China)

SSR-Encoder: Encoding Selective Subject Representation for Subject-Driven Generation 8069
Yuxuan Zhang (Shanghai Jiao Tong University), Yiren Song (National University of Singapore), Jiaming Liu (Xiaohongshu Inc.), Rui Wang (Beijing University of Posts and Telecommunications), Jinpeng Yu (ShanghaiTech University), Hao Tang (Carnegie Mellon University), Huaxia Li (Xiaohongshu Inc.), Xu Tang (Xiaohongshu Inc.), Yao Hu (Xiaohongshu Inc.), Han Pan (Shanghai Jiao Tong University), and Zhongliang Jing (Shanghai Jiao Tong University)

PEEKABOO: Interactive Video Generation via Masked-Diffusion 8079
Yash Jain (Microsoft), Anshul Nasery (University of Washington), Vibhav Vineet (Microsoft), and Harkirat Behl (Microsoft)

CoDeF: Content Deformation Fields for Temporally Consistent Video Processing 8089
Hao Ouyang (Hong Kong University of Science and Technology), Qiuyu Wang (Ant Group), Yuxi Xiao (Zhejiang University), Qingyan Bai (Hong Kong University of Science and Technology), Juntao Zhang (Hong Kong University of Science and Technology), Kecheng Zheng (Ant Group), Xiaowei Zhou (Zhejiang University), Qifeng Chen (Hong Kong University of Science and Technology), and Yujun Shen (Ant Group)

DreamMatcher: Appearance Matching Self-Attention for Semantically-Consistent Text-to-Image Personalization 8100
Jisu Nam (Korea University, Korea), Heesu Kim (NAVER Cloud, Korea), DongJae Lee (NAVER Cloud, Korea), Siyoon Jin (Korea University, Korea), Seungryong Kim (Korea University, Korea), and Seunggyu Chang (NAVER Cloud, Korea)

DreamComposer: Controllable 3D Object Generation via Multi-View Conditions 8111
Yunhan Yang (The University of Hong Kong), Yukun Huang (The University of Hong Kong), Xiaoyang Wu (The University of Hong Kong), Yuan-Chen Guo (VAST, Tsinghua University), Song-Hai Zhang (Tsinghua University), Hengshuang Zhao (The University of Hong Kong), Tong He (Shanghai Artificial Intelligence Lab), and Xihui Liu (The University of Hong Kong)

Shadow Generation for Composite Image Using Diffusion Model 8121
Qingyang Liu (Shanghai Jiao Tong University, China), Junqi You (Shanghai Jiao Tong University, China), Jianting Wang (Shanghai Jiao Tong University, China), Xinhao Tao (Shanghai Jiao Tong University, China), Bo Zhang (Shanghai Jiao Tong University, China), and Li Niu (Shanghai Jiao Tong University, China)

Adversarial Score Distillation: When score distillation meets GAN 8131
Min Wei (Beijing University of Posts and Telecommunications), Jingkai Zhou (Independent Researcher), Junyao Sun (South China University of Technology), and Xuesong Zhang (Beijing University of Posts and Telecommunications)

Uncertainty-Aware Source-Free Adaptive Image Super-Resolution with Wavelet Augmentation Transformer 8142
Yuang Ai (Institute of Automation, Chinese Academy of Sciences, China), Xiaoqiang Zhou (University of Science and Technology of China, China), Huaibo Huang (Institute of Automation, Chinese Academy of Sciences, China), Lei Zhang (The Hong Kong Polytechnic University, China), and Ran He (Institute of Automation, Chinese Academy of Sciences, China)

Animate Anyone: Consistent and Controllable Image-to-Video Synthesis for Character Animation 8153
Li Hu (Alibaba Group)

Person in Place: Generating Associative Skeleton-Guidance Maps for Human-Object Interaction Image Editing 8164
ChangHee Yang (Sogang University, AI Lab, CTO Division, LG Electronics), ChanHee Kang (Sogang University), Kyeongbo Kong (Pusan National University), Hanni Oh (Sogang University), and Suk-Ju Kang (Sogang University)

StableVITON: Learning Semantic Correspondence with Latent Diffusion Model for Virtual Try-On 8176
Jeongho Kim (KAIST, South Korea), Guojung Gu (KAIST, South Korea), Minho Park (KAIST, South Korea), Sunghyun Park (KAIST, South Korea), and Jaegul Choo (KAIST, South Korea)

Personalized Residuals for Concept-Driven Text-to-Image Generation 8186
Cusuh Ham (Georgia Institute of Technology, USA), Matthew Fisher (Adobe Research, USA), James Hays (Georgia Institute of Technology, USA), Nicholas Kolkin (Adobe Research, USA), Yuchen Liu (Adobe Research, USA), Richard Zhang (Adobe Research, USA), and Tobias Hinz (Adobe Research, USA)

UFOGen: You Forward Once Large Scale Text-to-Image Generation via Diffusion GANs 8196
Yanwu Xu (Boston University), Yang Zhao (Google), Zhisheng Xiao (Google), and Tingbo Hou (Google)

FlowVid: Taming Imperfect Optical Flows for Consistent Video-to-Video Synthesis 8207
Feng Liang (University of Texas at Austin, USA), Bichen Wu (Meta, USA), Jialiang Wang (Meta, USA), Licheng Yu (Meta, USA), Kunpeng Li (Meta, USA), Yinan Zhao (Meta, USA), Ishan Misra (Meta, USA), Jia-Bin Huang (Meta, USA), Peizhao Zhang (Meta, USA), Peter Vajda (Meta, USA), and Diana Marculescu (University of Texas at Austin, USA)

Readout Guidance: Learning Control from Diffusion Features 8217
Grace Luo (Google Research, UC Berkeley), Trevor Darrell (UC Berkeley), Oliver Wang (Google Research), Dan B Goldman (Google Research), and Aleksander Holynski (Google Research, UC Berkeley)

Diffusion Model Alignment Using Direct Preference Optimization 8228
Bram Wallace (OpenAI, USA), Meihua Dang (Stanford University, USA), Rafael Rafailov (Stanford University, USA), Linqi Zhou (Stanford University, USA), Aaron Lou (Stanford University, USA), Senthil Purushwalkam (Salesforce AI, USA), Stefano Ermon (Stanford University, USA), Caiming Xiong (Salesforce AI, USA), Shafiq Joty (Salesforce AI, USA), and Nikhil Naik (Salesforce AI, USA)

Diffusion Models Without Attention 8239
Jing Nathan Yan (Cornell University), Jiatao Gu (Apple), and Alexander M. Rush (Cornell University)

CommonCanvas: Open Diffusion Models Trained on Creative-Commons Images 8250
Aaron Gokaslan (Cornell), A. Feder Cooper (Cornell University; GenLaw), Jasmine Collins (Mosaic Research Databricks), Landan Seguin (Mosaic Research Databricks), Austin Jacobson (Mosaic Research Databricks), Mihir Patel (Mosaic Research Databricks), Jonathan Frankle (Mosaic Research Databricks), Cory Stephenson (Mosaic Research Databricks), and Volodymyr Kuleshov (Cornell)

Fairy: Fast Parallelized Instruction-Guided Video-to-Video Synthesis 8261
Bichen Wu (Meta), Ching-Yao Chuang (Meta), Xiaoyan Wang (Meta), Yichen Jia (Meta), Kapil Krishnakumar (Meta), Tong Xiao (Meta), Feng Liang (Meta), Licheng Yu (Meta), and Peter Vajda (Meta)

Edit One for All: Interactive Batch Image Editing 8271
Thao Nguyen (University of Wisconsin–Madison, USA), Utkarsh Ojha (University of Wisconsin–Madison, USA), Yuheng Li (University of Wisconsin–Madison, USA), Haotian Liu (University of Wisconsin–Madison, USA), and Yong Jae Lee (University of Wisconsin–Madison, USA)

Wavelet-based Fourier Information Interaction with Frequency Diffusion Adjustment for Underwater Image Restoration 8281
Chen Zhao (Nanjing Normal University, China), Weiling Cai (Nanjing Normal University, China), Chenyu Dong (Nanjing Normal University, China), and Chengwei Hu (Nanjing Normal University, China)

Accelerating Diffusion Sampling with Optimized Time Steps 8292
Shuchen Xue (Academy of Mathematics and Systems Science, Chinese Academy of Sciences), Zhaoqiang Liu (University of Electronic Science and Technology of China), Fei Chen (Huawei Noah's Ark Lab), Shifeng Zhang (Huawei Noah's Ark Lab), Tianyang Hu (Huawei Noah's Ark Lab), Enze Xie (Huawei Noah's Ark Lab), and Zhenguo Li (Huawei Noah's Ark Lab)

One-Shot Structure-Aware Stylized Image Synthesis 8302
Hansam Cho (Korea University), Jonghyun Lee (Korea University), Seunggyu Chang (NAVER Cloud), and Yonghyun Jeong (NAVER Cloud)

Selectively Informative Description can Reduce Undesired Embedding Entanglements in Text-to-Image Personalization 8312
Jimyeong Kim (Seoul National University), Jungwon Park (Seoul National University), and Wonjong Rhee (Seoul National University)

Observation-Guided Diffusion Probabilistic Models 8323
Junoh Kang (Seoul National University, Republic of Korea), Jinyoung Choi (Seoul National University, Republic of Korea), Sungik Choi (LG AI Research, Republic of Korea), and Bohyung Han (Seoul National University, Republic of Korea)

Scaling Up Video Summarization Pretraining with Large Language Models 8332
Dawit Mureja Argaw (KAIST), Seunghyun Yoon (Adobe Research), Fabian Caba Heilbron (Adobe Research), Hanieh Deilamsalehy (Adobe Research), Trung Bui (Adobe Research), Zhaowen Wang (Adobe Research), Franck Dernoncourt (Adobe Research), and Joon Son Chung (KAIST)

DREAM: Diffusion Rectification and Estimation-Adaptive Models 8342
Jinxin Zhou (Ohio State University, USA), Tianyu Ding (Microsoft Corporation, USA), Tianyi Chen (Microsoft Corporation, USA), Jiachen Jiang (Ohio State University, USA), Ilya Zharkov (Microsoft Corporation, USA), Zhihui Zhu (Ohio State University, USA), and Luming Liang (Microsoft Corporation, USA)

Clockwork Diffusion: Efficient Generation With Model-Step Distillation 8352
Amirhossein Habibian (Qualcomm AI Research), Amir Ghodrati (Qualcomm AI Research), Noor Fathima (Qualcomm AI Research), Guillaume Sautiere (Qualcomm AI Research), Risheek Garrepalli (Qualcomm AI Research), Fatih Porikli (Qualcomm AI Research), and Jens Petersen (Qualcomm AI Research)

SmartEdit: Exploring Complex Instruction-based Image Editing with Multimodal Large Language Models 8362
Yuzhou Huang (The Chinese University of Hong Kong, Shenzhen (CUHK-SZ); ARC Lab, Tencent PCG), Liangbin Xie (ARC Lab, Tencent PCG; University of Macau; Shenzhen Institute of Advanced Technology), Xintao Wang (ARC Lab, Tencent PCG; Tencent AI Lab), Ziyang Yuan (ARC Lab, Tencent PCG; Tsinghua University), Xiaodong Cun (Tencent AI Lab), Yixiao Ge (ARC Lab, Tencent PCG; Tencent AI Lab), Jiantao Zhou (University of Macau), Chao Dong (Shenzhen Institute of Advanced Technology; Shanghai Artificial Intelligence Laboratory), Rui Huang (The Chinese University of Hong Kong, Shenzhen (CUHK-SZ)), Ruimao Zhang (The Chinese University of Hong Kong, Shenzhen (CUHK-SZ)), and Ying Shan (ARC Lab, Tencent PCG; Tencent AI Lab)

CAT-DM: Controllable Accelerated Virtual Try-on with Diffusion Model 8372
Jianhao Zeng (Tianjin University, China), Dan Song (Tianjin University, China), Weizhi Nie (Tianjin University, China), Hongshuo Tian (Tianjin University, China), Tongtong Wang (Tencent LightSpeed Studio, China), and An-An Liu (Tianjin University, China)

Exact Fusion via Feature Distribution Matching for Few-shot Image Generation 8383
Yingbo Zhou (East China Normal University, China), Yutong Ye (East China Normal University, China), Pengyu Zhang (East China Normal University, China), Xian Wei (East China Normal University, China), and Mingsong Chen (East China Normal University, China)

Cross Initialization for Face Personalization of Text-to-Image Models 8393
Lianyu Pang (Sun Yat-sen University, China), Jian Yin (Sun Yat-sen University, China), Haoran Xie (Lingnan University, Hong Kong), Qiping Wang (East China Normal University, China), Qing Li (The Hong Kong Polytechnic University, Hong Kong), and Xudong Mao (Sun Yat-sen University, China)

EasyDrag: Efficient Point-based Manipulation on Diffusion Models 8404
Xingzhong Hou (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences; School of Computer Science and Technology, University of Chinese Academy of Sciences), Boxiao Liu (SenseTime Research), Yi Zhang (SenseTime Research), Jihao Liu (SenseTime Research; CUHK MMLab), Yu Liu (SenseTime Research), and Haihang You (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences)

MicroCinema: A Divide-and-Conquer Approach for Text-to-Video Generation 8414
Yanhui Wang (University of Science and Technology of China, China), Jianmin Bao (Microsoft Research Asia, China), Wenming Weng (University of Science and Technology of China, China), Ruoyu Feng (University of Science and Technology of China, China), Dacheng Yin (University of Science and Technology of China, China), Tao Yang (Xi'an Jiaotong University, China), Jingxu Zhang (University of Science and Technology of China, China), Qi Dai (Microsoft Research Asia, China), Zhiyuan Zhao (Microsoft Research Asia, China), Chunyu Wang (Microsoft Research Asia, China), Kai Qiu (Microsoft Research Asia, China), Yuhui Yuan (Microsoft Research Asia, China), Xiaoyan Sun (University of Science and Technology of China, China), Chong Luo (Microsoft Research Asia, China), and Baining Guo (Microsoft Research Asia, China)

Towards Memorization-Free Diffusion Models 8425
Chen Chen (University of Sydney), Daochang Liu (University of Sydney), and Chang Xu (University of Sydney)

SD-DiT: Unleashing the Power of Self-supervised Discrimination in Diffusion Transformer 8435
Rui Zhu (The Chinese University of HongKong, Shenzhen), Yingwei Pan (HiDream.ai Inc.), Yehao Li (HiDream.ai Inc.), Ting Yao (HiDream.ai Inc.), Zhenglong Sun (The Chinese University of HongKong, Shenzhen), Tao Mei (HiDream.ai Inc.), and Chang Wen Chen (The Hong Kong Polytechnic University)

Towards Effective Usage of Human-Centric Priors in Diffusion Models for Text-based Human Image Generation 8446
Junyan Wang (University of New South Wales, Australia), Zhenhong Sun (Australian National University, Australia), Zhiyu Tan (INF Technology, China), Xuanbai Chen (Carnegie Mellon University, USA), Weihua Chen (Alibaba DAMO Academy, China), Hao Li (Fudan University, China), Cheng Zhang (Carnegie Mellon University, USA), and Yang Song (University of New South Wales, Australia)

Text2QR: Harmonizing Aesthetic Customization and Scanning Robustness for Text-Guided QR Code Generation 8456
Guangyang Wu (Shanghai Jiao Tong University, China), Xiaohong Liu (Shanghai Jiao Tong University, China), Jun Jia (Shanghai Jiao Tong University), Xuehao Cui (Shanghai Jiao Tong University), and Guangtao Zhai (Shanghai Jiao Tong University)

Space-Time Diffusion Features for Zero-Shot Text-Driven Motion Transfer 8466
Danah Yatim (Weizmann Institute of Science, Israel), Rafail Fridman (Weizmann Institute of Science, Israel), Omer Bar-Tal (Weizmann Institute of Science, Israel), Yoni Kasten (NVIDIA Research, Israel), and Tali Dekel (Weizmann Institute of Science, Israel)

Video Frame Interpolation via Direct Synthesis with the Event-based Reference 8477
Yuhan Liu (Beijing University of Technology), Yongjian Deng (Beijing University of Technology), Hao Chen (Southeast University), and Zhen Yang (Beijing University of Technology)

DiffEditor: Boosting Accuracy and Flexibility on Diffusion-based Image Editing 8488
Chong Mou (Peking University, China), Xintao Wang (Tencent, China), Jiechong Song (Peking University, China), Ying Shan (Tencent, China), and Jian Zhang (Peking University, China)

EMOPortraits: Emotion-enhanced Multimodal One-shot Head Avatars 8498
Nikita Drobyshev (Imperial College London), Antoni Bigata Casademunt (Imperial College London), Konstantinos Vougioukas (Imperial College London), Zoe Landgraf (Imperial College London), Stavros Petridis (Imperial College London), and Maja Pantic (Imperial College London)

Spacetime Gaussian Feature Splatting for Real-Time Dynamic View Synthesis 8508
Zhan Li (OPPO US Research Center, USA; Portland State University, USA), Zhang Chen (OPPO US Research Center, USA), Zhong Li (OPPO US Research Center, USA), and Yi Xu (OPPO US Research Center, USA)

HOIDiffusion: Generating Realistic 3D Hand-Object Interaction Data 8521
Mengqi Zhang (University of California San Diego, USA), Yang Fu (University of California San Diego, USA), Zheng Ding (University of California San Diego, USA), Sifei Liu (NVIDIA, USA), Zhuowen Tu (University of California San Diego, USA), and Xiaolong Wang (University of California San Diego, USA)

Learned Representation-Guided Diffusion Models for Large-Image Generation 8532
Alexandros Graikos (Stony Brook University), Srikar Yellapragada (Stony Brook University), Minh-Quan Le (Stony Brook University), Saarthak Kapse (Stony Brook University), Prateek Prasanna (Stony Brook University), Joel Saltz (Stony Brook University), and Dimitris Samaras (Stony Brook University)

InstantBooth: Personalized Text-to-Image Generation without Test-Time Finetuning 8543
Jing Shi (Adobe Research), Wei Xiong (Adobe Inc.), Zhe Lin (Adobe Research), and Hyun Joon Jung (Adobe Inc.)

TokenCompose: Text-to-Image Diffusion with Token-level Supervision 8553
Zirui Wang (Princeton University), Zhizhou Sha (Tsinghua University), Zheng Ding (University of California, San Diego), Yilin Wang (Tsinghua University), and Zhuowen Tu (University of California, San Diego)

Geometry Transfer for Stylizing Radiance Fields 8565
Hyunyoung Jung (Seoul National University), Seonghyeon Nam (Meta Reality Labs), Nikolaos Sarafianos (Meta Reality Labs), Sungjoo Yoo (Seoul National University), Alexander Sorkine-Hornung (Meta Reality Labs), and Rakesh Ranjan (Meta Reality Labs)

Align Your Gaussians: Text-to-4D with Dynamic 3D Gaussians and Composed Diffusion Models 8576
Huan Ling (NVIDIA, Canada; Vector Institute, Canada; University of Toronto, Canada), Seung Wook Kim (NVIDIA, Canada; Vector Institute, Canada; University of Toronto, Canada), Antonio Torralba (MIT, USA), Sanja Fidler (NVIDIA, Canada; Vector Institute, Canada; University of Toronto, Canada), and Karsten Kreis (NVIDIA, Canada)

DreamSalon: A Staged Diffusion Framework for Preserving Identity-Context in Editable Face Generation 8589
Haonan Lin (Xi'an Jiaotong University)

Video-P2P: Video Editing with Cross-attention Control 8599
Shaoteng Liu (The Chinese University of Hong Kong), Yuechen Zhang (The Chinese University of Hong Kong), Wenbo Li (The Chinese University of Hong Kong), Zhe Lin (Adobe), and Jiaya Jia (The Chinese University of Hong Kong)

PAIR Diffusion: A Comprehensive Multimodal Object-Level Image Editor 8609
Vidit Goel (Picsart AI Research, USA; SHI Labs @ Georgia Tech & UIUC, USA), Elia Peruzzo (University of Trento, Italy), Yifan Jiang (University of Texas Austin, USA), Dejia Xu (University of Texas Austin, USA), Xingqian Xu (Picsart AI Research, USA; SHI Labs @ Georgia Tech & UIUC, USA), Nicu Sebe (University of Trento, Italy), Trevor Darrell (University of California Berkeley, USA), Zhangyang Wang (Picsart AI Research, USA; University of Texas Austin, USA), and Humphrey Shi (Picsart AI Research, USA; SHI Labs @ Georgia Tech & UIUC, USA)

ArtAdapter: Text-to-Image Style Transfer using Multi-Level Style Encoder and Explicit Adaptation 8619
Dar-Yen Chen (PicCollage, Taiwan), Hamish Tennent (PicCollage, Taiwan), and Ching-Wen Hsu (PicCollage, Taiwan)

DemoCaricature: Democratising Caricature Generation with a Rough Sketch 8629
Dar-Yen Chen (University of Surrey, United Kingdom), Ayan Kumar Bhunia (University of Surrey, United Kingdom), Subhadeep Koley (University of Surrey, United Kingdom), Aneeshan Sain (University of Surrey, United Kingdom), Pinaki Nath Chowdhury (University of Surrey, United Kingdom), and Yi-Zhe Song (University of Surrey, United Kingdom)

PhotoMaker: Customizing Realistic Human Photos via Stacked ID Embedding 8640
Zhen Li (Nankai University, China), Mingdeng Cao (Tokyo University, Japan), Xintao Wang (Tencent, China), Zhongang Qi (Tencent, China), Ming-Ming Cheng (Nankai University, China), and Ying Shan (Tencent, China)

Predicated Diffusion: Predicate Logic-Based Attention Guidance for Text-to-Image Diffusion Models 8651
Kota Sueyoshi (Osaka University) and Takashi Matsubara (Osaka University)

SNED: Superposition Network Architecture Search for Efficient Video Diffusion Model 8661
Zhengang Li (Northeastern University, USA), Yan Kang (Adobe Research, USA), Yuchen Liu (Adobe Research, USA), Difan Liu (Adobe Research, USA), Tobias Hinz (Adobe, USA), Feng Liu (Adobe Research, USA), and Yanzhi Wang (Northeastern University, USA)

TRIP: Temporal Residual Learning with Image Noise Prior for Image-to-Video Diffusion Models 8671
Zhongwei Zhang (University of Science and Technology of China, Hefei, China), Fuchen Long (HiDream.ai Inc.), Yingwei Pan (HiDream.ai Inc.), Zhaofan Qiu (HiDream.ai Inc.), Ting Yao (HiDream.ai Inc.), Yang Cao (University of Science and Technology of China, Hefei, China), and Tao Mei (HiDream.ai Inc.)

Prompt-Free Diffusion: Taking "Text" out of Text-to-Image Diffusion Models 8682
Xingqian Xu (University of Illinois at Urbana-Champaign; Picsart AI Research), Jiayi Guo (Tsinghua University), Zhangyang Wang (UT Austin; Picsart AI Research), Gao Huang (Tsinghua University), Irfan Essa (Georgia Tech), and Humphrey Shi (Georgia Tech; Picsart AI Research)

DEADiff: An Efficient Stylization Diffusion Model with Disentangled Representations 8693
Tianhao Qi (University of Science and Technology of China), Shancheng Fang (University of Science and Technology of China), Yanze Wu (ByteDance Inc.), Hongtao Xie (University of Science and Technology of China), Jiawei Liu (ByteDance Inc.), Lang Chen (ByteDance Inc.), Qian He (ByteDance Inc.), and Yongdong Zhang (University of Science and Technology of China)

FRESCO: Spatial-Temporal Correspondence for Zero-Shot Video Translation 8703
Shuai Yang (Peking University, China), Yifan Zhou (Nanyang Technological University, Singapore), Ziwei Liu (Nanyang Technological University, Singapore), and Chen Change Loy (Nanyang Technological University, Singapore)

Correcting Diffusion Generation through Resampling 8713
Yujian Liu (UC Santa Barbara, USA), Yang Zhang (MIT-IBM Watson AI Lab, USA), Tommi Jaakkola (MIT CSAIL, USA), and Shiyu Chang (UC Santa Barbara, USA)

AnyScene: Customized Image Synthesis with Composited Foreground 8724
Ruidong Chen (Tianjin University), Lanjun Wang (Tianjin University), Weizhi Nie (Tianjin University), Yongdong Zhang (University of Science and Technology of China), and An-An Liu (Tianjin University)

Grid Diffusion Models for Text-to-Video Generation 8734
Taegyeong Lee (Ulsan National Institute of Science and Technology), Soyeong Kwon (Ulsan National Institute of Science and Technology), and Taehwan Kim (Ulsan National Institute of Science and Technology)

Direct2.5: Diverse Text-to-3D Generation via Multi-view 2.5D Diffusion 8744
Yuanxun Lu (Nanjing University), Jingyang Zhang (Apple), Shiwei Li (Apple), Tian Fang (Apple), David McKinnon (Apple), Yanghai Tsin (Apple), Long Quan (The Hong Kong University of Science and Technology), Xun Cao (Nanjing University), and Yao Yao (Nanjing University)

Anomaly Score: Evaluating Generative Models and Individual Generated Images based on Complexity and Vulnerability 8754
Jaehui Hwang (Yonsei University, Korea), Junghyuk Lee (Yonsei University, Korea), and Jong-Seok Lee (Yonsei University, Korea)

Zero-Painter: Training-Free Layout Control for Text-to-Image Synthesis 8764
Marianna Ohanyan (Picsart AI Research (PAIR)), Hayk Manukyan (Picsart AI Research (PAIR)), Zhangyang Wang (Picsart AI Research (PAIR), UT Austin), Shant Navasardyan (Picsart AI Research (PAIR)), and Humphrey Shi (Picsart AI Research (PAIR), Georgia Tech)

X-Adapter: Adding Universal Compatibility of Plugins for Upgraded Diffusion Model 8775
Lingmin Ran (National University of Singapore, Singapore), Xiaodong Cun (Tencent, China), Jia-Wei Liu (National University of Singapore, Singapore), Rui Zhao (National University of Singapore, Singapore), Song Zijie (Fudan University, China), Xintao Wang (Tencent, China), Jussi Keppo (National University of Singapore, Singapore), and Mike Zheng Shou (National University of Singapore, Singapore)

Neural Point Cloud Diffusion for Disentangled 3D Shape and Appearance Generation 8785
Philipp Schröppel (University of Freiburg, Germany), Christopher Wewer (Max Planck Institute for Informatics, Saarland Informatics Campus, Germany), Jan Eric Lenssen (Max Planck Institute for Informatics, Saarland Informatics Campus, Germany), Eddy Ilg (Saarland University, Saarland Informatics Campus, Germany), and Thomas Brox (University of Freiburg, Germany)

Style Injection in Diffusion: A Training-free Approach for Adapting Large-scale Diffusion Models for Style Transfer 8795
Jiwoo Chung (Sungkyunkwan University), Sangeek Hyun (Sungkyunkwan University), and Jae-Pil Heo (Sungkyunkwan University)

Vlogger: Make Your Dream A Vlog 8806
Shaobin Zhuang (Shanghai Jiao Tong University), Kunchang Li (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences), Xinyuan Chen (Shanghai AI Laboratory), Yaohui Wang (Shanghai AI Laboratory), Ziwei Liu (S-Lab, Nanyang Technological University), Yu Qiao (Shanghai AI Laboratory), and Yali Wang (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences)

Faces that Speak: Jointly Synthesising Talking Face and Speech from Text 8818
Youngjoon Jang (KAIST, Republic of Korea), Ji-Hoon Kim (KAIST, Republic of Korea), Junseok Ahn (KAIST, Republic of Korea), Doyeop Kwak (KAIST, Republic of Korea), Hong-Sun Yang (42dot, Republic of Korea), Yoon-Cheol Ju (42dot, Republic of Korea), Il-Hwan Kim (42dot, Republic of Korea), Byeong-Yeol Kim (42dot, Republic of Korea), and Joon Son Chung (KAIST, Republic of Korea)

Prompt Augmentation for Self-supervised Text-guided Image Manipulation 8829
Rumeysa Bodur (Imperial College London, UK), Binod Bhattarai (University of Aberdeen, UK), and Tae-Kyun Kim (KAIST, South Korea)

DragDiffusion: Harnessing Diffusion Models for Interactive Point-based Image Editing 8839
Yujun Shi (National University of Singapore), Chuhui Xue (ByteDance Inc.), Jun Hao Liew (ByteDance Inc.), Jiachun Pan (National University of Singapore), Hanshu Yan (ByteDance Inc.), Wenqing Zhang (ByteDance Inc.), Vincent Y. F. Tan (National University of Singapore), and Song Bai (ByteDance Inc.)

Make Pixels Dance: High-Dynamic Video Generation 8850
Yan Zeng (ByteDance Research, China), Guoqiang Wei (ByteDance Research, China), Jiani Zheng (ByteDance Research), Jiaxin Zou (ByteDance Research), Yang Wei (ByteDance Research), Yuchen Zhang (ByteDance Research), and Hang Li (ByteDance Research)

LEDITS++: Limitless Image Editing using Text-to-Image Models 8861
Manuel Brack (TU Darmstadt; DFKI), Felix Friedrich (TU Darmstadt; hessian.AI), Katharia Kornmeier (TU Darmstadt), Linoy Tsaban (Huggingface), Patrick Schramowski (TU Darmstadt; DFKI; hessian.AI), Kristian Kersting (TU Darmstadt; DFKI; hessian.AI), and Apolinario Passos (Huggingface)

Emu Edit: Precise Image Editing via Recognition and Generation Tasks 8871
Shelly Sheynin (GenAI, Meta), Adam Polyak (GenAI, Meta), Uriel Singer (GenAI, Meta), Yuval Kirstain (GenAI, Meta), Amit Zohar (GenAI, Meta), Oron Ashual (GenAI, Meta), Devi Parikh (GenAI, Meta), and Yaniv Taigman (GenAI, Meta)

Concept Weaver: Enabling Multi-Concept Fusion in Text-to-Image Models 8880
Gihyun Kwon (KAIST, Korea), Simon Jenni (Adobe, USA), Dingzeyu Li (Adobe, USA), Joon-Young Lee (Adobe, USA), Jong Chul Ye (KAIST, Korea), and Fabian Caba Heilbron (Adobe, USA)

ACT-Diffusion: Efficient Adversarial Consistency Training for One-step Diffusion Models 8890
Fei Kong (University of Electronic Science and Technology of China), Jinhao Duan (Drexel University), Lichao Sun (Lehigh University), Hao Cheng (The Hong Kong University of Science and Technology (Guangzhou)), Renjing Xu (The Hong Kong University of Science and Technology (Guangzhou)), Hengtao Shen (University of Electronic Science and Technology of China), Xiaofeng Zhu (University of Electronic Science and Technology of China), Xiaoshuang Shi (University of Electronic Science and Technology of China), and Kaidi Xu (Drexel University)

3D Geometry-Aware Deformable Gaussian Splatting for Dynamic View Synthesis 8900
Zhicheng Lu (Northwestern Polytechnical University, China), Xiang Guo (Northwestern Polytechnical University, China), Le Hui (Northwestern Polytechnical University, China), Tianrui Chen (Northwestern Polytechnical University, China), Min Yang (Samsung R&D Institute, China), Xiao Tang (Samsung R&D Institute, China), Feng Zhu (Samsung R&D Institute, China), and Yuchao Dai (Northwestern Polytechnical University, China)

Boosting Diffusion Models with Moving Average Sampling in Frequency Domain 8911
Yurui Qian (University of Science and Technology of China), Qi Cai (HiDream.ai Inc.), Yingwei Pan (HiDream.ai Inc.), Yehao Li (HiDream.ai Inc.), Ting Yao (HiDream.ai Inc.), Qibin Sun (University of Science and Technology of China), and Tao Mei (HiDream.ai Inc.)

NoiseCollage: A Layout-Aware Text-to-Image Diffusion Model Based on Noise Cropping and Merging 8921
Takahiro Shirakawa (Kyushu University, Japan) and Seiichi Uchida (Kyushu University, Japan)

NeRF On-the-go: Exploiting Uncertainty for Distractor-free NeRFs in the Wild 8931
Weining Ren (ETH Zurich), Zihan Zhu (ETH Zurich), Boyang Sun (ETH Zurich), Jiaqi Chen (ETH Zurich), Marc Pollefeys (ETH Zurich; Microsoft), and Songyou Peng (ETH Zurich; MPI for Intelligent Systems, Tubingen)

Using Human Feedback to Fine-tune Diffusion Models without Any Reward Model 8941
Kai Yang (Tsinghua University), Jian Tao (Tsinghua University), Jiafei Lyu (Tsinghua University), Chunjiang Ge (Tsinghua University), Jiaxin Chen (Parametrix Technology Company Ltd.), Weihan Shen (Parametrix Technology Company Ltd.), Xiaolong Zhu (Parametrix Technology Company Ltd.), and Xiu Li (Tsinghua University)

GeneAvatar: Generic Expression-Aware Volumetric Head Avatar Editing from a Single Image 8952
Chong Bao (Zhejiang University), Yinda Zhang (Google), Yuan Li (Zhejiang University), Xiyu Zhang (Zhejiang University), Bangbang Yang (ByteDance), Hujun Bao (Zhejiang University), Marc Pollefeys (ETH Zurich), Guofeng Zhang (Zhejiang University), and Zhaopeng Cui (Zhejiang University)

MaskPLAN: Masked Generative Layout Planning from Partial Input 8964
Hang Zhang (Digital Building Technologies, ETH Zurich), Anton Savov (Digital Building Technologies, ETH Zurich), and Benjamin Dillenburger (Digital Building Technologies, ETH Zurich)

WOUAF: Weight Modulation for User Attribution and Fingerprinting in Text-to-Image Diffusion Models 8974
Changhoon Kim (Arizona State University), Kyle Min (Intel Labs), Maitreya Patel (Arizona State University), Sheng Cheng (Arizona State University), and Yezhou Yang (Arizona State University)

Transcending Forgery Specificity with Latent Space Augmentation for Generalizable Deepfake Detection 8984
Zhiyuan Yan (The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen), China), Yuhao Luo (The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen), China), Siwei Lyu (University at Buffalo, State University of New York, USA), Qingshan Liu (Nanjing University of Information Science and Technology, China), and Baoyuan Wu (The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen), China)

SCEdit: Efficient and Controllable Image Diffusion Generation via Skip Connection Editing 8995
Zeyinzi Jiang (Alibaba Group, China), Chaojie Mao (Alibaba Group, China), Yulin Pan (Alibaba Group, China), Zhen Han (Alibaba Group, China), and Jingfeng Zhang (Alibaba Group, China)

CONFORM: Contrast is All You Need for High-Fidelity Text-to-Image Diffusion Models 9005
Tuna Han Salih Meral (Virginia Tech, USA), Enis Simsar (ETH Zürich, Switzerland), Federico Tombari (Google, Switzerland), and Pinar Yanardag (Virginia Tech, USA)

TI2V-Zero: Zero-Shot Image Conditioning for Text-to-Video Diffusion Models 9015
Haomiao Ni (The Pennsylvania State University, USA), Bernhard Egger (Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany), Suhas Lohit (Mitsubishi Electric Research Laboratories, USA), Anoop Cherian (Mitsubishi Electric Research Laboratories, USA), Ye Wang (Mitsubishi Electric Research Laboratories, USA), Toshiaki Koike-Akino (Mitsubishi Electric Research Laboratories, USA), Sharon X. Huang (The Pennsylvania State University, USA), and Tim K. Marks (Mitsubishi Electric Research Laboratories, USA)

HIVE: Harnessing Human Feedback for Instructional Visual Editing 9026
Shu Zhang (Salesforce AI Research), Xinyi Yang (Salesforce AI Research), Yihao Feng (Salesforce AI Research), Can Qin (Salesforce AI Research), Chia-Chih Chen (Nvidia), Ning Yu (Salesforce AI Research), Zeyuan Chen (Salesforce AI Research), Huan Wang (Salesforce AI Research), Silvio Savarese (Salesforce AI Research), Stefano Ermon (Stanford University), Caiming Xiong (Salesforce AI Research), and Ran Xu (Salesforce AI Research)

Taming Mode Collapse in Score Distillation for Text-to-3D Generation 9037
Peihao Wang (The University of Texas at Austin), Dejia Xu (The University of Texas at Austin), Zhiwen Fan (The University of Texas at Austin), Dilin Wang (Meta Reality Labs), Sreyas Mohan (Meta Reality Labs), Forrest Iandola (Meta Reality Labs), Rakesh Ranjan (Meta Reality Labs), Yilei Li (Meta Reality Labs), Qiang Liu (The University of Texas at Austin), Zhangyang Wang (The University of Texas at Austin), and Vikas Chandra (Meta Reality Labs)

CoDi: Conditional Diffusion Distillation for Higher-Fidelity and Faster Image Generation 9048
Kangfu Mei (Johns Hopkins University), Mauricio Delbracio (Google Research), Hossein Talebi (Google Research), Zhengzhong Tu (Google Research), Vishal M. Patel (Johns Hopkins University), and Peyman Milanfar (Google Research)

Universal Robustness via Median Randomized Smoothing for Real-World Super-Resolution 9059
Zakariya Chaouai (Paris-Saclay University, France) and Mohamed Tamaazousti (Paris-Saclay University, France)

ECLIPSE: A Resource-Efficient Text-to-Image Prior for Image Generations 9069
Maitreya Patel (Arizona State University, USA), Changhoon Kim (Arizona State University, USA), Sheng Cheng (Arizona State University, USA), Chitta Baral (Arizona State University, USA), and Yezhou Yang (Arizona State University, USA)

CAMEL: CAusal Motion Enhancement Tailored for Lifting Text-driven Video Editing 9079
Guiwei Zhang (Beihang University), Tianyu Zhang (Du Xiaoman Financial), Guanglin Niu (Beihang University), Zichang Tan (Baidu Inc), Yalong Bai (Du Xiaoman Financial), and Qing Yang (Du Xiaoman Financial)

FreeCustom: Tuning-Free Customized Image Generation for Multi-Concept Composition 9089
Ganggui Ding (Zhejiang University, China), Canyu Zhao (Zhejiang University), Wen Wang (Zhejiang University, China), Zhen Yang (Zhejiang University, China), Zide Liu (Zhejiang University, China), Hao Chen (Zhejiang University, China), and Chunhua Shen (Zhejiang University, China)

Amodal Completion via Progressive Mixed Context Diffusion 9099
Katherine Xu (University of Pennsylvania), Lingzhi Zhang (Adobe Inc.), and Jianbo Shi (University of Pennsylvania)

Named Entity Driven Zero-Shot Image Manipulation 9110
Zhida Feng (Wuhan University of Science and Technology.; Baidu Inc.), Li Chen (Wuhan University of Science and Technology), Jing Tian (National University of Singapore), JiaXiang Liu (Baidu Inc.), and Shikun Feng (Baidu Inc.)

Learning Degradation-unaware Representation with Prior-based Latent Transformations for Blind Face Restoration 9120
Lianxin Xie (South China University of Technology), Csbingbing Zheng (South China University of Technology), Wen Xue (South China University of Technology), Le Jiang (South China University of Technology), Cheng Liu (Shantou University), Si Wu (South China University of Technology), and Hau San Wong (City University of Hong Kong)

AEROBLADE: Training-Free Detection of Latent Diffusion Images Using Autoencoder Reconstruction Error 9130
Jonas Ricker (Ruhr University Bochum, Germany), Denis Lukovnikov (Ruhr University Bochum, Germany), and Asja Fischer (Ruhr University Bochum, Germany)

VRetouchEr: Learning Cross-frame Feature Interdependence with Imperfection Flow for Face Retouching in Videos 9141
Wen Xue (South China University of Technology), Le Jiang (South China University of Technology), Lianxin Xie (South China University of Technology), Si Wu (South China University of Technology), Yong Xu (South China University of Technology), and Hau San Wong (City University of Hong Kong)

Generative Unlearning for Any Identity 9151
Juwon Seo (Kyung Hee University, Republic of Korea), Sung-Hoon Lee (Kyung Hee University, Republic of Korea), Tae-Young Lee (Kyung Hee University, Republic of Korea), Seungjun Moon (KLleon Tech, Republic of Korea), and Gyeong-Moon Park (Kyung Hee University, Republic of Korea)

Doubly Abductive Counterfactual Inference for Text-based Image Editing 9162
Xue Song (Fudan University, China), Jiequan Cui (Nanyang Technological University, Singapore), Hanwang Zhang (Skywork AI, Singapore; Nanyang Technological University, Singapore), Jingjing Chen (Fudan University, China), Richang Hong (Hefei University of Technology, China), and Yu-Gang Jiang (Fudan University, China)

Text-conditional Attribute Alignment across Latent Spaces for 3D Controllable Face Image Synthesis 9172
Feifan Xu (South China University of Technology,China), Rui Li (Shantou University,China), Si Wu (South China University of Technology,China), Yong Xu (South China University of Technology,China), and Hau San Wong (City University of Hong Kong,China)

Customization Assistant for Text-to-Image Generation 9182
Yufan Zhou (Adobe Research), Ruiyi Zhang (Adobe Research), Jiuxiang Gu (Adobe Research), and Tong Sun (Adobe Research)

Contrastive Denoising Score for Text-guided Latent Diffusion Image Editing 9192
Hyelin Nam (KAIST), Gihyun Kwon (KAIST), Geon Yeong Park (KAIST), and Jong Chul Ye (KAIST)

Arbitrary-Scale Image Generation and Upsampling using Latent Diffusion Model and Implicit Neural Decoder 9202
Jinseok Kim (KAIST) and Tae-Kyun Kim (KAIST)

VMC: Video Motion Customization using Temporal Attention Adaption for Text-to-Video Diffusion Models 9212
Hyeonho Jeong (Korea Advanced Institute of Science and Technology, Republic of Korea), Geon Yeong Park (Korea Advanced Institute of Science and Technology, Republic of Korea), and Jong Chul Ye (Korea Advanced Institute of Science and Technology, Republic of Korea)

Visual Layout Composer: Image-Vector Dual Diffusion Model for Design Layout Generation 9222
Mohammad Amin Shabani (Simon Fraser University), Zhaowen Wang (Adobe Research), Difan Liu (Adobe Research), Nanxuan Zhao (Adobe Research), Jimei Yang (Adobe Research), and Yasutaka Furukawa (Simon Fraser University)

Learning Spatial Adaptation and Temporal Coherence in Diffusion Models for Video Super-Resolution 9232
Zhikai Chen (University of Science and Technology of China), Fuchen Long (HiDream.ai Inc.), Zhaofan Qiu (HiDream.ai Inc.), Ting Yao (HiDream.ai Inc.), Wengang Zhou (University of Science and Technology of China), Jiebo Luo (University of Rochester), and Tao Mei (HiDream.ai Inc.)

Open-Vocabulary Attention Maps with Token Optimization for Semantic Segmentation in Diffusion Models 9242
Pablo Marcos-Manchón (University of Barcelona, Spain), Roberto Alcover-Couso (Autonomous University of Madrid, Spain), Juan C. SanMiguel (Autonomous University of Madrid, Spain), and José M. Martínez (Autonomous University of Madrid, Spain)

Combining Frame and GOP Embeddings for Neural Video Representation 9253
Jens Eirik Saethre (DisneyResearch|Studios, Switzerland; ETH Zürich, Switzerland), Roberto Azevedo (DisneyResearch|Studios, Switzerland), and Christopher Schroers (DisneyResearch|Studios, Switzerland)

PLACE: Adaptive Layout-Semantic Fusion for Semantic Image Synthesis 9264
Zhengyao Lv (The University of Hong Kong, China), Yuxiang Wei (Harbin Institute of Technology, China), Wangmeng Zuo (Harbin Institute of Technology, China), and Kwan-Yee K. Wong (The University of Hong Kong, China)

Your Student is Better Than Expected: Adaptive Teacher-Student Collaboration for Text-Conditional Diffusion Models 9275
Nikita Starodubcev (Yandex), Dmitry Baranchuk (Yandex), Artem Fedorov (Yandex), and Artem Babenko (Yandex)

Mitigating Motion Blur in Neural Radiance Fields with Events and Frames 9286
Marco Cannici (University of Zurich, Switzerland) and Davide Scaramuzza (University of Zurich, Switzerland)

Unmixing Before Fusion: A Generalized Paradigm for Multi-Source-based Hyperspectral Image Synthesis 9297
Yang Yu (Wuhan University), Erting Pan (Wuhan University), Xinya Wang (Wuhan University), Yuheng Wu (Wuhan University), Xiaoguang Mei (Wuhan University), and Jiayi Ma (Wuhan University)

Rethinking FID: Towards a Better Evaluation Metric for Image Generation 9307
Sadeep Jayasumana (Google Research), Srikumar Ramalingam (Google Research), Andreas Veit (Google Research), Daniel Glasner (Google Research), Ayan Chakrabarti (Google Research), and Sanjiv Kumar (Google Research)

MarkovGen: Structured Prediction for Efficient Text-to-Image Generation 9316
Sadeep Jayasumana (Google Research), Daniel Glasner (Google Research), Srikumar Ramalingam (Google Research), Andreas Veit (Google Research), Ayan Chakrabarti (Google Research), and Sanjiv Kumar (Google Research)

DisCo: Disentangled Control for Realistic Human Dance Generation 9326
Tan Wang (Nanyang Technological University), Linjie Li (Microsoft), Kevin Lin (Microsoft), Yuanhao Zhai (University at Buffalo), Chung-Ching Lin (Microsoft), Zhengyuan Yang (Microsoft), Hanwang Zhang (Nanyang Technological University), Zicheng Liu (Advanced Micro Devices), and Lijuan Wang (Microsoft)

The Devil is in the Details: StyleFeatureEditor for Detail-Rich StyleGAN Inversion and High Quality Image Editing 9337
Denis Bobkov (HSE University), Vadim Titov (AIRI), Aibek Alanov (AIRI; HSE University), and Dmitry Vetrov (Constructor University, Bremen)

C3: High-Performance and Low-Complexity Neural Compression from a Single Image or Video . 9347
Hyunjik Kim (Google DeepMind), Matthias Bauer (Google DeepMind), Lucas Theis (Google DeepMind), Jonathan Richard Schwarz (Harvard University), and Emilien Dupont (Google DeepMind)

LightIt: Illumination Modeling and Control for Diffusion Models 9359
Peter Kocsis (Technical Univesrity of Munich, Germany), Julien Philip (Adobe Research, United Kingdom), Kalyan Sunkavalli (Adobe Research, USA), Matthias Nießner (Technical Univesrity of Munich, Germany), and Yannick Hold-Geoffroy (Adobe Research, USA)

Rethinking the Spatial Inconsistency in Classifier-Free Diffusion Guidance 9370
Dazhong Shen (Shanghai Artificial Intelligence Laboratory), Guanglu Song (SenseTime Research), Zeyue Xue (The University of Hong Kong), Fu-Yun Wang (The Chinese University of Hong Kong), and Yu Liu (Shanghai Artificial Intelligence Laboratory)

InitNO: Boosting Text-to-Image Diffusion Models via Initial Noise Optimization 9380
Xiefan Guo (Beihang University, China), Jinlin Liu (Individual, China), Miaomiao Cui (Individual, China), Jiankai Li (Beihang University, China), Hongyu Yang (Beihang University, China), and Di Huang (Beihang University, China)

On the Diversity and Realism of Distilled Dataset: An Efficient Dataset Distillation Paradigm 9390
Peng Sun (Zhejiang University; Westlake University), Bei Shi (Westlake University), Daiwei Yu (Independent Researcher), and Tao Lin (Westlake University)

On the Scalability of Diffusion-based Text-to-Image Generation 9400
Hao Li (Amazon, USA), Yang Zou (Amazon, USA), Ying Wang (Amazon, USA), Orchid Majumder (Amazon, USA), Yusheng Xie (Amazon, USA), R. Manmatha (Amazon, USA), Ashwin Swaminathan (Amazon, USA), Zhuowen Tu (Amazon, USA), Stefano Ermon (Amazon, USA), and Stefano Soatto (Amazon, USA)

Distilling ODE Solvers of Diffusion Models into Smaller Steps 9410
Sanghwan Kim (ETH Zürich, Switzerland), Hao Tang (ETH Zürich, Switzerland), and Fisher Yu (ETH Zürich, Switzerland)

ZeroNVS: Zero-Shot 360-Degree View Synthesis from a Single Image 9420
Kyle Sargent (Stanford University), Zizhang Li (Stanford University), Tanmay Shah (Google), Charles Herrmann (Google), Hong-Xing Yu (Stanford University), Yunzhi Zhang (Stanford University), Eric Ryan Chan (Stanford University), Dmitry Lagun (Google), Li Fei-Fei (Stanford University), Deqing Sun (Google), and Jiajun Wu (Stanford University)

Fixed Point Diffusion Models 9430
Xingjian Bai (Oxford University) and Luke Melas-Kyriazi (Oxford University)

Gaussian Shell Maps for Efficient 3D Human Generation 9441
Rameen Abdal (Stanford University, USA), Wang Yifan (Stanford University, USA), Zifan Shi (HKUST, Hong Kong), Yinghao Xu (Stanford University, USA), Ryan Po (Stanford University, USA), Zhengfei Kuang (Stanford University, USA), Qifeng Chen (HKUST, Hong Kong), Dit-Yan Yeung (HKUST, Hong Kong), and Gordon Wetzstein (Stanford University, USA)

Inversion-Free Image Editing with Language-Guided Diffusion Models 9452
Sihan Xu (University of Michigan, Ann Arbor), Yidong Huang (University of Michigan, Ann Arbor), Jiayi Pan (University of California, Berkeley), Ziqiao Ma (University of Michigan, Ann Arbor), and Joyce Chai (University of Michigan, Ann Arbor)

TIGER: Time-Varying Denoising Model for 3D Point Cloud Generation with Diffusion Process ... 9462
Zhiyuan Ren (Michigan State University), Minchul Kim (Michigan State University), Feng Liu (Michigan State University), and Xiaoming Liu (Michigan State University)

Beyond First-Order Tweedie: Solving Inverse Problems using Latent Diffusion 9472
Litu Rout (The University of Texas at Austin, USA), Yujia Chen (Google, USA), Abhishek Kumar (Google Deepmind), Constantine Caramanis (The University of Texas at Austin, USA), Sanjay Shakkottai (The University of Texas at Austin, USA), and Wen-Sheng Chu (Google, USA)

U-VAP: User-specified Visual Appearance Personalization via Decoupled Self Augmentation 9482
You Wu (Institute of Computing Technology, CAS, China), Kean Liu (Institute of Computing Technology, CAS, China), Xiaoyue Mi (Institute of Computing Technology, CAS, China), Fan Tang (Institute of Computing Technology, CAS, China), Juan Cao (Institute of Computing Technology, CAS, China), and Jintao Li (Institute of Computing Technology, CAS, China)

Orals 3A 3D from Single View

Repurposing Diffusion-Based Image Generators for Monocular Depth Estimation 9492
Bingxin Ke (ETH Zürich, Switzerland), Anton Obukhov (ETH Zürich, Switzerland), Shengyu Huang (ETH Zürich, Switzerland), Nando Metzger (ETH Zürich, Switzerland), Rodrigo Caye Daudt (ETH Zürich, Switzerland), and Konrad Schindler (ETH Zürich, Switzerland)

EscherNet: A Generative Model for Scalable View Synthesis 9503
Xin Kong (Dyson Robotics Lab, Imperial College London), Shikun Liu (Dyson Robotics Lab, Imperial College London), Xiaoyang Lyu (The University of Hong Kong), Marwan Taher (Dyson Robotics Lab, Imperial College London), Xiaojuan Qi (The University of Hong Kong), and Andrew J. Davison (Dyson Robotics Lab, Imperial College London)

WALT3D: Generating Realistic Training Data from Time-Lapse Imagery for Reconstructing Dynamic Objects Under Occlusion 9514
Khiem Vuong (Carnegie Mellon University, USA), N Dinesh Reddy (Amazon, USA), Robert Tamburo (Carnegie Mellon University, USA), and Srinivasa G. Narasimhan (Carnegie Mellon University, USA)

Diffusion-FOF: Single-View Clothed Human Reconstruction via Diffusion-Based Fourier Occupancy Field 9525
Yuanzhen Li (Wuhan University), Fei Luo (Wuhan University), and Chunxia Xiao (Wuhan University)

Rethinking Inductive Biases for Surface Normal Estimation 9535
Gwangbin Bae (Dyson Robotics Lab, Imperial College London) and Andrew J. Davison (Dyson Robotics Lab, Imperial College London)

Orals 3B Vision, Language, and Reasoning

Comparing the Decision-Making Mechanisms by Transformers and CNNs via Explanation Methods .. 9546
Mingqi Jiang (Oregon State University, USA), Saeed Khorram (Oregon State University, USA), and Li Fuxin (Oregon State University, USA)

MMMU: A Massive Multi-discipline Multimodal Understanding and Reasoning Benchmark for Expert AGI 9556
Xiang Yue (IN.AI Research, USA), Yuansheng Ni (University of Waterloo, Canada), Kai Zhang (The Ohio State University, USA), Tianyu Zheng (Independent Researcher), Ruoqi Liu (The Ohio State University, USA), Ge Zhang (University of Waterloo, Canada), Samuel Stevens (The Ohio State University, USA), Dongfu Jiang (University of Waterloo, Canada), Weiming Ren (University of Waterloo, Canada), Yuxuan Sun (Independent Researcher), Cong Wei (University of Waterloo, Canada), Botao Yu (The Ohio State University, USA), Ruibin Yuan (Carnegie Mellon University, USA), Renliang Sun (University of Waterloo, Canada), Ming Yin (Princeton University, USA), Boyuan Zheng (The Ohio State University, USA), Zhenzhu Yang (Independent Researcher), Yibo Liu (University of Victoria, Canada), Wenhao Huang (Independent Researcher), Huan Sun (The Ohio State University, USA), Yu Su (The Ohio State University, USA), and Wenhu Chen (University of Waterloo, Canada)

Eyes Wide Shut? Exploring the Visual Shortcomings of Multimodal LLMs 9568
Shengbang Tong (New York University), Zhuang Liu (FAIR, Meta), Yuexiang Zhai (University of California, Berkeley), Yi Ma (University of California, Berkeley), Yann LeCun (New York University), and Saining Xie (New York University)

LISA: Reasoning Segmentation via Large Language Model 9579
Xin Lai (The Chinese University of Hong Kong, China), Zhuotao Tian (Harbin Institute of Technology (Shenzhen), China), Yukang Chen (The Chinese University of Hong Kong, China), Yanwei Li (The Chinese University of Hong Kong, China), Yuhui Yuan (Microsoft Research Asia, China), Shu Liu (SmartMore, China), and Jiaya Jia (The Chinese University of Hong Kong, China)

Visual Program Distillation: Distilling Tools and Programmatic Reasoning into Vision-Language Models 9590
Yushi Hu (University of Washington), Otilia Stretcu (Google Research), Chun-Ta Lu (Google Research), Krishnamurthy Viswanathan (Google Research), Kenji Hata (Google Research), Enming Luo (Google Research), Ranjay Krishna (University of Washington), and Ariel Fuxman (Google Research)

Orals 3C Medical and Physics-based Vision

EventPS: Real-Time Photometric Stereo Using an Event Camera 9602
Bohan Yu (Peking University, China), Jieji Ren (Shanghai Jiao Tong University, China), Jin Han (The University of Tokyo, Japan), Feishi Wang (Peking University, China), Jinxiu Liang (Peking University, China), and Boxin Shi (Peking University, China)

EvDiG: Event-guided Direct and Global Components Separation 9612
Xinyu Zhou (Peking University, China), Peiqi Duan (Peking University, China), Boyu Li (Peking University, China), Chu Zhou (Peking University, China), Chao Xu (Peking University, China), and Boxin Shi (Peking University, China)

MemSAM: Taming Segment Anything Model for Echocardiography Video Segmentation 9622
Xiaolong Deng (Shenzhen University, China), Huisi Wu (Shenzhen University, China), Runhao Zeng (Shenzhen University, China), and Jing Qin (The Hong Kong Polytechnic University, Hong Kong)

Transcriptomics-guided Slide Representation Learning in Computational Pathology 9632
Guillaume Jaume (Harvard Medical School, USA), Lukas Oldenburg (Mass General Brigham, USA), Anurag Vaidya (Harvard Medical School, USA), Richard J. Chen (Harvard Medical School, USA), Drew F.K. Williamson (Mass General Brigham, USA), Thomas Peeters (Mass General Brigham, USA), Andrew H. Song (Harvard Medical School, USA), and Faisal Mahmood (Harvard Medical School, USA)

Correlation-aware Coarse-to-fine MLPs for Deformable Medical Image Registration 9645
Mingyuan Meng (The University of Sydney, Australia), Dagan Feng (The University of Sydney, Australia), Lei Bi (Shanghai Jiao Tong University, China), and Jinman Kim (The University of Sydney, Australia)

Poster Session 3

G3DR: Generative 3D Reconstruction in ImageNet 9655
Pradyumna Reddy (Huawei Noah's Ark Lab UK), Ismail Elezi (Huawei Noah's Ark Lab UK), and Jiankang Deng (Huawei Noah's Ark Lab UK)

CityDreamer: Compositional Generative Model of Unbounded 3D Cities 9666
Haozhe Xie (Nanyang Technological University), Zhaoxi Chen (Nanyang Technological University), Fangzhou Hong (Nanyang Technological University), and Ziwei Liu (Nanyang Technological University)

6D-Diff: A Keypoint Diffusion Framework for 6D Object Pose Estimation 9676
Li Xu (Singapore University of Technology and Design), Haoxuan Qu (Singapore University of Technology and Design), Yujun Cai (Nanyang Technological University), and Jun Liu (Singapore University of Technology and Design)

Generative Proxemics: A Prior for 3D Social Interaction from Images 9687
Lea Müller (UC Berkeley, USA), Vickie Ye (UC Berkeley, USA), Georgios Pavlakos (UT Austin, USA), Michael Black (MPI for Intelligent Systems, Tübingen, Germany), and Angjoo Kanazawa (UC Berkeley, USA)

MVD-Fusion: Single-view 3D via Depth-consistent Multi-view Generation 9698
Hanzhe Hu (Carnegie Mellon University), Zhizhuo Zhou (Stanford University), Varun Jampani (Stability AI), and Shubham Tulsiani (Carnegie Mellon University)

WorDepth: Variational Language Prior for Monocular Depth Estimation 9708
Ziyao Zeng (Yale University, USA), Daniel Wang (Yale University, USA), Fengyu Yang (Yale University, USA), Hyoungseob Park (Yale University, USA), Stefano Soatto (University of California, Los Angeles, USA), Dong Lao (University of California, Los Angeles, USA), and Alex Wong (Yale University, USA)

Free3D: Consistent Novel View Synthesis without 3D Representation 9720
Chuanxia Zheng (University of Oxford) and Andrea Vedaldi (University of Oxford)

PostureHMR: Posture Transformation for 3D Human Mesh Recovery 9732
Yu-Pei Song (Southwest Jiaotong University, China; Engineering Research Center of Sustainable Urban Intelligent Transportation, China), Xiao Wu (Southwest Jiaotong University, China; Engineering Research Center of Sustainable Urban Intelligent Transportation, China), Zhaoquan Yuan (Southwest Jiaotong University, China; Engineering Research Center of Sustainable Urban Intelligent Transportation, China), Jian-Jun Qiao (Southwest Jiaotong University, China; Engineering Research Center of Sustainable Urban Intelligent Transportation, China), and Qiang Peng (Southwest Jiaotong University, China; Engineering Research Center of Sustainable Urban Intelligent Transportation, China)

3DFIRES: Few Image 3D REconstruction for Scenes with Hidden Surfaces 9742
Linyi Jin (University of Michigan), Nilesh Kulkarni (University of Michigan), and David F. Fouhey (New York University)

Learning the 3D Fauna of the Web 9752
Zizhang Li (Stanford University, USA), Dor Litvak (University of Texas at Austin, USA), Ruining Li (University of Oxford, United Kingdom), Yunzhi Zhang (Stanford University, USA), Tomas Jakab (University of Oxford, United Kingdom), Christian Rupprecht (University of Oxford, United Kingdom), Shangzhe Wu (Stanford University, USA), Andrea Vedaldi (University of Oxford, United Kingdom), and Jiajun Wu (Stanford University, USA)

Bilateral Propagation Network for Depth Completion 9763
Jie Tang (National University of Defense Technology, China), Fei-Peng Tian (LightIllusions, China), Boshi An (Peking University, China), Jian Li (National University of Defense Technology, China), and Ping Tan (The Hong Kong University of Science and Technology, China)

Efficient Privacy-Preserving Visual Localization Using 3D Ray Clouds 9773
Heejoon Moon (Hanyang University, South Korea), Chunghwan Lee (Hanyang University, South Korea), and Je Hyeong Hong (Hanyang University, South Korea)

EpiDiff: Enhancing Multi-View Synthesis via Localized Epipolar-Constrained Diffusion 9784
Zehuan Huang (Beihang University, China), Hao Wen (Beihang University, China), Junting Dong (Shanghai AI Laboratory, China), Yaohui Wang (Shanghai AI Laboratory, China), Yangguang Li (VAST, China), Xinyuan Chen (Shanghai AI Laboratory, China), Yan-Pei Cao (VAST, China), Ding Liang (VAST, China), Yu Qiao (Shanghai AI Laboratory, China), Bo Dai (Shanghai AI Laboratory, China), and Lu Sheng (Beihang University, China)

Doodle Your 3D: From Abstract Freehand Sketches to Precise 3D Shapes 9795
Hmrishav Bandyopadhyay (University of Surrey), Subhadeep Koley (University of Surrey, iFlyTek-Surrey Joint Research Centre on Artificial Intelligence), Ayan Das (University of Surrey), Ayan Kumar Bhunia (University of Surrey), Aneeshan Sain (University of Surrey), Pinaki Nath Chowdhury (University of Surrey), Tao Xiang (University of Surrey, iFlyTek-Surrey Joint Research Centre on Artificial Intelligence), and Yi-Zhe Song (University of Surrey, iFlyTek-Surrey Joint Research Centre on Artificial Intelligence)

LowRankOcc: Tensor Decomposition and Low-Rank Recovery for Vision-based 3D Semantic Occupancy Prediction 9806
Linqing Zhao (Tsinghua University), Xiuwei Xu (Tsinghua University), Ziwei Wang (Carnegie Mellon University), Yunpeng Zhang (PhiGent Robotics), Borui Zhang (Tsinghua University), Wenzhao Zheng (University of California, Berkeley), Dalong Du (PhiGent Robotics), Jie Zhou (Tsinghua University), and Jiwen Lu (Tsinghua University)

CNC-Net: Self-Supervised Learning for CNC Machining Operations 9816
Mohsen Yavartanoo (Seoul National University), Sangmin Hong (Seoul National University), Reyhaneh Neshatavar (Seoul National University), and Kyoung Mu Lee (Seoul National University)

Reconstructing Hands in 3D with Transformers 9826
Georgios Pavlakos (UT Austin), Dandan Shan (University of Michigan), Ilija Radosavovic (UC Berkeley), Angjoo Kanazawa (UC Berkeley), David Fouhey (New York University), and Jitendra Malik (UC Berkeley)

Boosting Self-Supervision for Single-View Scene Completion via Knowledge Distillation 9837
Keonhee Han (Technical University of Munich, Germany), Dominik Muhle (Technical University of Munich, Germany), Felix Wimbauer (Technical University of Munich, Germany), and Daniel Cremers (Technical University of Munich, Germany)

Know Your Neighbors: Improving Single-View Reconstruction via Spatial Vision-Language Reasoning 9848
Rui Li (ETH Zürich, Switzerland), Tobias Fischer (ETH Zürich, Switzerland), Mattia Segu (ETH Zürich, Switzerland), Marc Pollefeys (ETH Zürich, Switzerland), Luc Van Gool (ETH Zürich, Switzerland), and Federico Tombari (Google, Switzerland; Technical University of Munich, Germany)

Depth Prompting for Sensor-Agnostic Depth Estimation 9859
Jin-Hwi Park (Gwangju Institute of Science and Technology (GIST)), Chanhwi Jeong (Gwangju Institute of Science and Technology (GIST)), Junoh Lee (Gwangju Institute of Science and Technology (GIST)), and Hae-Gon Jeon (Gwangju Institute of Science and Technology (GIST))

ViewFusion: Towards Multi-View Consistency via Interpolated Denoising 9870
Xianghui Yang (The University of Sydney, Australia), Yan Zuo (Amazon, Australia), Sameera Ramasinghe (Amazon, Australia), Loris Bazzani (Amazon, Germany), Gil Avraham (Amazon, Australia), and Anton van den Hengel (The University of Adelaide, Australia)

Slice3D: Multi-Slice, Occlusion-Revealing, Single View 3D Reconstruction 9881
Yizhi Wang (Simon Fraser University), Wallace Lira (Simon Fraser University), Wenqi Wang (Tsinghua University), Ali Mahdavi-Amiri (Simon Fraser University), and Hao Zhang (Simon Fraser University)

Consistent3D: Towards Consistent High-Fidelity Text-to-3D Generation with Deterministic Sampling Prior 9892
Zike Wu (Nanyang Technological University, Singapore), Pan Zhou (Singapore Management University, Singapore), Xuanyu Yi (Nanyang Technological University, Singapore), Xiaoding Yuan (Johns Hopkins University, USA), and Hanwang Zhang (Nanyang Technological University, Singapore)

GigaPose: Fast and Robust Novel Object Pose Estimation via One Correspondence 9903
Van Nguyen Nguyen (Ecole des Ponts ParisTech), Thibault Groueix (Adobe, USA), Mathieu Salzmann (EPFL), and Vincent Lepetit (Ecole des Ponts ParisTech)

RichDreamer: A Generalizable Normal-Depth Diffusion Model for Detail Richness in Text-to-3D 9914
Lingteng Qiu (SSE, CUHKSZ; Alibaba Group), Guanying Chen (SSE, CUHKSZ; FNii CUHKSZ), Xiaodong Gu (Alibaba Group), Qi Zuo (Alibaba Group), Mutian Xu (SSE, CUHKSZ), Yushuang Wu (SSE, CUHKSZ, FNii CUHKSZ), Weihao Yuan (Alibaba Group), Zilong Dong (Alibaba Group), Liefeng Bo (Alibaba Group), and Xiaoguang Han (SSE, CUHKSZ; FNii, CUHKSZ)

Elite360D: Towards Efficient 360 Depth Estimation via Semantic- and Distance-Aware Bi-Projection Fusion 9926
Hao Ai (Hong Kong University of Science and Technology (Guangzhou Campus), China) and Lin Wang (Hong Kong University of Science and Technology (Guangzhou Campus), China and Hong Kong University of Science and Technology, Hongkong)

SIFU: Side-view Conditioned Implicit Function for Real-world Usable Clothed Human Reconstruction 9936
Zechuan Zhang (Zhejiang University, China), Zongxin Yang (Zhejiang University, China), and Yi Yang (Zhejiang University, China)

Diffusion Time-step Curriculum for One Image to 3D Generation 9948
Xuanyu Yi (Nanyang Technological University, Singapore), Zike Wu (Nanyang Technological University, Singapore), Qingshan Xu (Nanyang Technological University, Singapore), Pan Zhou (Singapore Management University, Singapore), Joo-Hwee Lim (Institute for Infocomm Research, Singapore), and Hanwang Zhang (Nanyang Technological University, Singapore)

SecondPose: SE(3)-Consistent Dual-Stream Feature Fusion for Category-Level Pose Estimation.... 9959
Yamei Chen (Technical University of Munich, Germany), Yan Di (Technical University of Munich, Germany), Guangyao Zhai (Technical University of Munich, Germany & Munich Center for Machine Learning, Germany), Fabian Manhardt (Google, Germany), Chenyangguang Zhang (Tsinghua University, China), Ruida Zhang (Tsinghua University, China), Federico Tombari (Technical University of Munich, Germany & Google, Switzerland), Nassir Navab (Technical University of Munich, Germany), and Benjamin Busam (Technical University of Munich, Germany & Munich Center for Machine Learning, Germany & 3dwe.ai, Germany)

Wonder3D: Single Image to 3D using Cross-Domain Diffusion 9970
Xiaoxiao Long (The University of Hong Kong), Yuan-Chen Guo (Tsinghua University), Cheng Lin (The University of Hong Kong), Yuan Liu (The University of Hong Kong), Zhiyang Dou (The University of Hong Kong), Lingjie Liu (University of Pennsylvania), Yuexin Ma (Shanghai Tech University), Song-Hai Zhang (Tsinghua University), Marc Habermann (MPI Informatik), Christian Theobalt (MPI Informatik), and Wenping Wang (Texas A&M University)

En3D: An Enhanced Generative Model for Sculpting 3D Humans from 2D Synthetic Data 9981
Yifang Men (Alibaba Group), Biwen Lei (Alibaba Group), Yuan Yao (Alibaba Group), Miaomiao Cui (Alibaba Group), Zhouhui Lian (Peking University), and Xuansong Xie (Alibaba Group)

MOHO: Learning Single-view Hand-held Object Reconstruction with Multi-view Occlusion-Aware Supervision 9992
Chenyangguang Zhang (Tsinghua University, China), Guanlong Jiao (Tsinghua University, China), Yan Di (Technical University of Munich, Germany), Gu Wang (Tsinghua University, China), Ziqin Huang (Tsinghua University, China), Ruida Zhang (Tsinghua University, China), Fabian Manhardt (Google, Switzerland), Bowen Fu (Tsinghua University, China), Federico Tombari (Technical University of Munich, Germany; Google, Switzerland), and Xiangyang Ji (Tsinghua University, China)

Template Free Reconstruction of Human-object Interaction with Procedural Interaction Generation 10003
Xianghui Xie (University of Tübingen; Max Planck Institute for Informatics), Bharat Lal Bhatnagar (Meta), Jan Eric Lenssen (Max Planck Institute for Informatics), and Gerard Pons-Moll (University of Tübingen; Max Planck Institute for Informatics)

PatchFusion: An End-to-End Tile-Based Framework for High-Resolution Monocular Metric Depth Estimation 10016
Zhenyu Li (King Abdullah University of Science and Technology (KAUST)), Shariq Farooq Bhat (King Abdullah University of Science and Technology (KAUST)), and Peter Wonka (King Abdullah University of Science and Technology (KAUST))

SPAD: Spatially Aware Multi-View Diffusers 10026
Yash Kant (University of Toronto), Aliaksandr Siarohin (Snap Research), Ziyi Wu (University of Toronto), Michael Vasilkovsky (Snap Research), Guocheng Qian (Snap Research), Jian Ren (Snap Research), Riza Alp Guler (Snap Research), Bernard Ghanem (KAUST), Sergey Tulyakov (Snap Research), and Igor Gilitschenski (University of Toronto)

GenFlow: Generalizable Recurrent Flow for 6D Pose Refinement of Novel Objects 10039
Sungphill Moon (NAVER LABS, South Korea), Hyeontae Son (NAVER LABS, South Korea), Dongcheol Hur (NAVER LABS, South Korea), and Sangwook Kim (NAVER LABS, South Korea)

PointInfinity: Resolution-Invariant Point Diffusion Models 10050
Zixuan Huang (FAIR at Meta, USA), Justin Johnson (FAIR at Meta, USA), Shoubhik Debnath (FAIR at Meta, USA), James M. Rehg (University of Illinois at Urbana-Champaign, USA), and Chao-Yuan Wu (FAIR at Meta, USA)

ZeroShape: Regression-based Zero-shot Shape Reconstruction 10061
Zixuan Huang (University of Illinois at Urbana-Champaign, USA), Stefan Stojanov (Georgia Institute of Technology, USA), Anh Thai (Georgia Institute of Technology, USA), Varun Jampani (Stability AI, USA), and James M. Rehg (University of Illinois at Urbana-Champaign, USA)

One-2-3-45++: Fast Single Image to 3D Objects with Consistent Multi-View Generation and 3D Diffusion 10072
Minghua Liu (UC San Diego, USA), Ruoxi Shi (UC San Diego, USA), Linghao Chen (UC San Diego, USA), Zhuoyang Zhang (Tsinghua University, China), Chao Xu (UCLA, USA), Xinyue Wei (UC San Diego, USA), Hansheng Chen (Stanford University), Chong Zeng (Zhejiang University, China), Jiayuan Gu (UC San Diego, USA), and Hao Su (UC San Diego, USA)

ConTex-Human: Free-View Rendering of Human from a Single Image with Texture-Consistent Synthesis 10084
Xiangjun Gao (The Hong Kong University of Science and Technology), Xiaoyu Li (Tencent AI Lab), Chaopeng Zhang (Tencent AI Lab), Qi Zhang (Tencent AI Lab), Yanpei Cao (Tencent AI Lab), Ying Shan (Tencent AI Lab), and Long Quan (The Hong Kong University of Science and Technology)

MatchU: Matching Unseen Objects for 6D Pose Estimation from RGB-D Images 10095
Junwen Huang (Technical University of Munich, Germany, Munich Center for Machine Learning, Germany), Hao Yu (Technical University of Munich, Germany), Kuan-Ting Yu (XYZ Robotics), Nassir Navab (Technical University of Munich, Germany, Munich Center for Machine Learning, Germany), Slobodan Ilic (Technical University of Munich, Germay), and Benjamin Busam (Technical University of Munich, Germany, Munich Center for Machine Learning, Germany, 3dwe.ai, Germany)

UniDepth: Universal Monocular Metric Depth Estimation 10106
Luigi Piccinelli (ETH Zurich), Yung-Hsu Yang (ETH Zurich), Christos Sakaridis (ETH Zurich), Mattia Segu (ETH Zurich), Siyuan Li (ETH Zurich), Luc Van Gool (ETH Zurich, INSAIT), and Fisher Yu (ETH Zurich)

G-NeRF: Geometry-enhanced Novel View Synthesis from Single-View Images 10117
Zixiong Huang (South China University of Technology, China), Qi Chen (The University of Adelaide, Australia), Libo Sun (The University of Adelaide, Australia), Yifan Yang (South China University of Technology, China), Naizhou Wang (Guangzhou Shiyuan Electronics Co., Ltd, China), Qi Wu (The University of Adelaide, Australia), and Mingkui Tan (South China University of Technology, China)

3DToonify: Creating Your High-Fidelity 3D Stylized Avatar Easily from 2D Portrait Images 10127
Yifang Men (Alibaba Group), Hanxi Liu (Peking University), Yuan Yao (Alibaba Group), Miaomiao Cui (Alibaba Group), Xuansong Xie (Alibaba Group), and Zhouhui Lian (Peking University)

Adaptive Fusion of Single-View and Multi-View Depth for Autonomous Driving 10138
Junda Cheng (Huazhong University of Science and Technology), Wei Yin (DJI Technology), Kaixuan Wang (DJI Technology), Xiaozhi Chen (DJI Technology), Shijie Wang (Huazhong University of Science and Technology), and Xin Yang (Huazhong University of Science and Technology)

HiPose: Hierarchical Binary Surface Encoding and Correspondence Pruning for RGB-D 6DoF Object Pose Estimation 10148
Yongliang Lin (Zhejiang University, China), Yongzhi Su (German Research Center for Artificial Intelligence (DFKI), Germany), Praveen Nathan (German Research Center for Artificial Intelligence (DFKI), Germany), Sandeep Inuganti (German Research Center for Artificial Intelligence (DFKI), Germany), Yan Di (Technische Universität München, Germany), Martin Sundermeyer (Google, Germany), Fabian Manhardt (Google, Germany), Didier Stricker (German Research Center for Artificial Intelligence (DFKI), Germany), Jason Rambach (German Research Center for Artificial Intelligence (DFKI), Germany), and Yu Zhang (Zhejiang University, China)

HandBooster: Boosting 3D Hand-Mesh Reconstruction by Conditional Synthesis and Sampling of Hand-Object Interactions 10159
Hao Xu (The Chinese University of Hong Kong), Haipeng Li (University of Electronic Science and Technology of China), Yinqiao Wang (The Chinese University of Hong Kong), Shuaicheng Liu (University of Electronic Science and Technology of China), and Chi-Wing Fu (The Chinese University of Hong Kong)

3D-SceneDreamer: Text-Driven 3D-Consistent Scene Generation 10170
Songchun Zhang (Zhejiang University, China), Yibo Zhang (Jilin University, China), Quan Zheng (Institute of Software, Chinese Academy of Sciences, China), Rui Ma (Jilin University, China), Wei Hua (Zhejiang Lab, China), Hujun Bao (Zhejiang University, China), Weiwei Xu (Zhejiang University, China), and Changqing Zou (Zhejiang University, Zhejiang Lab, China)

NViST: In the Wild New View Synthesis from a Single Image with Transformers 10181
Wonbong Jang (University College London) and Lourdes Agapito (University College London)

CAD: Photorealistic 3D Generation via Adversarial Distillation 10194
Ziyu Wan (City University of Hong Kong), Despoina Paschalidou (Stanford University), Ian Huang (Stanford University), Hongyu Liu (HKUST), Bokui Shen (Stanford University), Xiaoyu Xiang (Independent Researcher), Jing Liao (City University of Hong Kong), and Leonidas Guibas (Stanford University)

Splatter Image: Ultra-Fast Single-View 3D Reconstruction 10208
Stanislaw Szymanowicz (University of Oxford, United Kingdom), Chrisitian Rupprecht (University of Oxford, United Kingdom), and Andrea Vedaldi (University of Oxford, United Kingdom)

Joint Reconstruction of 3D Human and Object via Contact-Based Refinement Transformer 10218
Hyeongjin Nam (Seoul National University), Daniel Sungho Jung (Seoul National University), Gyeongsik Moon (Codec Avatars Lab, Meta), and Kyoung Mu Lee (Seoul National University)

Sculpt3D: Multi-View Consistent Text-to-3D Generation with Sparse 3D Prior 10228
*Cheng Chen (Nanyang Technological University, Singapore), Xiaofeng Yang (Nanyang Technological University, Singapore), Fan Yang (Nanyang Technological University, Singapore), Chengzeng Feng (Nanyang Technological University, Singapore), Zhoujie Fu (Nanyang Technological University, Singapore), Chuan-Sheng Foo (A*STAR, Singapore), Guosheng Lin (Nanyang Technological University, Singapore), and Fayao Liu (A*STAR, Singapore)*

Object Pose Estimation via the Aggregation of Diffusion Features 10238
Tianfu Wang (Shenyang Institute of Automation, China), Guosheng Hu (Oosto, UK), and Hongguang Wang (Shenyang Institute of Automation, China)

MonoCD: Monocular 3D Object Detection with Complementary Depths 10248
Longfei Yan (Huazhong University of Science and Technology, China), Pei Yan (Huazhong University of Science and Technology, China), Shengzhou Xiong (Huazhong University of Science and Technology, China), Xuanyu Xiang (Huazhong University of Science and Technology, China), and Yihua Tan (Huazhong University of Science and Technology, China)

MultiDiff: Consistent Novel View Synthesis from a Single Image 10258
Norman Müller (Meta Reality Labs, Switzerland), Katja Schwarz (Meta Reality Labs, Switzerland), Barbara Rössle (Technical University of Munich, Germany), Lorenzo Porzi (Meta Reality Labs, Switzerland), Samuel Rota Bulò (Meta Reality Labs, Switzerland), Matthias Nießner (Technical University of Munich, Germany), and Peter Kontschieder (Meta Reality Labs, Switzerland)

SeaBird: Segmentation in Bird's View with Dice Loss Improves Monocular 3D Detection of Large Objects 10269
Abhinav Kumar (Michigan State University, USA), Yuliang Guo (Bosch Research North America, USA), Xinyu Huang (Bosch Research North America, USA), Liu Ren (Bosch Research North America, USA), and Xiaoming Liu (Michigan State University, USA)

Learning Occupancy for Monocular 3D Object Detection 10281
Liang Peng (FABU Inc., China), Junkai Xu (Zhejiang University, China), Haoran Cheng (Zhejiang University, China), Zheng Yang (FABU Inc., China), Xiaopei Wu (Zhejiang University, China), Wei Qian (FABU Inc., China), Wenxiao Wang (Zhejiang University, China), Boxi Wu (Zhejiang University, China), and Deng Cai (Zhejiang University, China)

NeRFDeformer: NeRF Transformation from a Single View via 3D Scene Flows 10293
Zhenggang Tang (University of Illinois Urbana-Champaign, USA), Zhongzheng Ren (University of Illinois Urbana-Champaign, USA), Xiaoming Zhao (University of Illinois Urbana-Champaign, USA), Bowen Wen (NVIDIA, USA), Jonathan Tremblay (NVIDIA, USA), Stan Birchfield (NVIDIA, USA), and Alexander Schwing (University of Illinois Urbana-Champaign, USA)

R-Cyclic Diffuser: Reductive and Cyclic Latent Diffusion for 3D Clothed Human Digitalization 10304
*Kennard Yanting Chan (Nanyang Technological University, Singapore), Fayao Liu (A*STAR, Singapore), Guosheng Lin (Nanyang Technological University, Singapore), Chuan Sheng Foo (A*STAR, Singapore), and Weisi Lin (Nanyang Technological University, Singapore)*

Unleashing Network Potentials for Semantic Scene Completion 10314
Fengyun Wang (Nanjing University of Science and Technology), Qianru Sun (Singapore Management University), Dong Zhang (The Hong Kong University of Science and Technology), and Jinhui Tang (Nanjing University of Science and Technology)

Triplane Meets Gaussian Splatting: Fast and Generalizable Single-View 3D Reconstruction with Transformers 10324
Zi-Xin Zou (Tsinghua University, China), Zhipeng Yu (VAST, China), Yuan-Chen Guo (Tsinghua University, China), Yangguang Li (VAST, China), Ding Liang (VAST, China), Yan-Pei Cao (VAST, China), and Song-Hai Zhang (Qinghai Univeristy, China; Tsinghua University, China)

VOODOO 3D: Volumetric Portrait Disentanglement For One-Shot 3D Head Reenactment 10336
Phong Tran (Mohamed bin Zayed University of Artificial Intelligence), Egor Zakharov (ETH Zurich), Long-Nhat Ho (Mohamed bin Zayed University of Artificial Intelligence), Anh Tuan Tran (VinAI Research), Liwen Hu (Pinscreen), and Hao Li (Mohamed bin Zayed University of Artificial Intelligence, Pinscreen)

Compressed 3D Gaussian Splatting for Accelerated Novel View Synthesis 10349
Simon Niedermayr (Technical University of Munich, Germany), Josef Stumpfegger (Technical University of Munich, Germany), and Rüdiger Westermann (Technical University of Munich, Germany)

Morphable Diffusion: 3D-Consistent Diffusion for Single-image Avatar Creation 10359
Xiyi Chen (ETH Zürich, Switzerland), Marko Mihajlovic (ETH Zürich, Switzerland), Shaofei Wang (ETH Zürich, Switzerland; University of Tübingen, Germany; Tübingen AI Center, Germany), Sergey Prokudin (ETH Zürich, Switzerland; University Hospital Balgrist, Switzerland), and Siyu Tang (ETH Zürich, Switzerland)

Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data 10371
Lihe Yang (The University of Hong Kong, Hong Kong), Bingyi Kang (TikTok, Singapore), Zilong Huang (TikTok, Singapore), Xiaogang Xu (The Chinese University of Hong Kong, Hong Kong), Jiashi Feng (TikTok, Singapore), and Hengshuang Zhao (The University of Hong Kong, Hong Kong)

SAOR: Single-View Articulated Object Reconstruction 10382
Mehmet Aygun (University of Edinburgh) and Oisin Mac Aodha (University of Edinburgh)

HOISDF: Constraining 3D Hand-Object Pose Estimation with Global Signed Distance Fields 10392
Haozhe Qi (EPFL), Chen Zhao (EPFL), Mathieu Salzmann (EPFL), and Alexander Mathis (EPFL)

Diffusion-driven GAN Inversion for Multi-Modal Face Image Generation 10403
Jihyun Kim (Yonsei University, South Korea), Changjae Oh (Queen Mary University of London, UK), Hoseok Do (AI Lab, CTO Division, LG Electronics, South Korea), Soohyun Kim (AI Lab, CTO Division, LG Electronics, South Korea), and Kwanghoon Sohn (Yonsei University, South Korea)

Novel View Synthesis with View-Dependent Effects from a Single Image 10413
Juan Luis Gonzalez Bello (Korea Advanced Institute of Science and Technology, Republic of Korea) and Munchurl Kim (Korea Advanced Institute of Science and Technology, Republic of Korea)

Weakly-Supervised Emotion Transition Learning for Diverse 3D Co-speech Gesture Generation 10424
Xingqun Qi (The Hong Kong University of Science and Technology), Jiahao Pan (The Hong Kong University of Science and Technology), Peng Li (Hong Kong University of Science and Technology), Ruibin Yuan (Hong Kong University of Science and Technology), Xiaowei Chi (Hong Kong University of Science and Technology), Mengfei Li (Hong Kong University of Science and Technology), Wenhan Luo (Hong Kong University of Science and Technology), Wei Xue (Hong Kong University of Science and Technology), Shanghang Zhang (Peking University), Qifeng Liu (Hong Kong University of Science and Technology), and Yike Guo (Hong Kong University of Science and Technology)

Seg2Reg: Differentiable 2D Segmentation to 1D Regression Rendering for 360 Room Layout Reconstruction 10435
Cheng Sun (NVIDIA), Wei-En Tai (National Tsing Hua University, Taiwan), Yu-Lin Shih (National Tsing Hua University, Taiwan), Kuan-Wei Chen (National Tsing Hua University, Taiwan), Yong-Jing Syu (National Tsing Hua University, Taiwan), Kent Selwyn The (National Tsing Hua University, Taiwan), Yu-Chiang Frank Wang (National Taiwan University, Taiwan), and Hwann-Tzong Chen (National Tsing Hua University, Taiwan)

Mining Supervision for Dynamic Regions in Self-Supervised Monocular Depth Estimation 10446
Hoang Chuong Nguyen (Australian National University), Tianyu Wang (Australian National University), Jose M. Alvarez (NVIDIA), and Miaomiao Liu (Australian National University)

DiffPortrait3D: Controllable Diffusion for Zero-Shot Portrait View Synthesis 10456
Yuming Gu (University of Southern California, USA), Hongyi Xu (Bytedance, USA), You Xie (Bytedance, USA), Guoxian Song (Bytedance, USA), Yichun Shi (Bytedance, USA), Di Chang (University of Southern California, USA), Jing Yang (University of Southern California, USA), and Linjie Luo (Bytedance, USA)

3D-LFM: Lifting Foundation Model 10466
Mosam Dabhi (Carnegie Mellon University, USA), László A. Jeni (Carnegie Mellon University, USA), and Simon Lucey (The University of Adelaide, Australia)

MRC-Net: 6-DoF Pose Estimation with MultiScale Residual Correlation 10476
Yuelong Li (Amazon Inc.), Yafei Mao (Amazon Inc.), Raja Bala (Amazon Inc.), and Sunil Hadap (Amazon Inc.)

DiffusionGAN3D: Boosting Text-guided 3D Generation and Domain Adaptation by Combining 3D GANs and Diffusion Priors 10487
Biwen Lei (Alibaba Group), Kai Yu (Alibaba Group), Mengyang Feng (Alibaba Group), Miaomiao Cui (Alibaba Group), and Xuansong Xie (Alibaba Group)

VS: Reconstructing Clothed 3D Human from Single Image via Vertex Shift 10498
Leyuan Liu (Central China Normal University, China), Yuhan Li (Central China Normal University, China), Yunqi Gao (Central China Normal University, China), Changxin Gao (Huazhong University of Science and Technology, China), Yuanyuan Liu (China University of Geosciences (Wuhan), China), and Jingying Chen (Central China Normal University, China)

Weakly Supervised Monocular 3D Detection with a Single-View Image 10508
Xueying Jiang (Nanyang Technological University, Singapore), Sheng Jin (Nanyang Technological University, Singapore), Lewei Lu (Sensetime Research, China), Xiaoqin Zhang (Zhejiang University of Technology, China), and Shijian Lu (Nanyang Technological University, Singapore)

From-Ground-To-Objects: Coarse-to-Fine Self-supervised Monocular Depth Estimation of Dynamic Objects with Ground Contact Prior 10519
Jaeho Moon (Korea Advanced Institute of Science and Technology, Republic of Korea), Juan Luis Gonzalez Bello (Korea Advanced Institute of Science and Technology, Republic of Korea), Byeongjun Kwon (Korea Advanced Institute of Science and Technology, Republic of Korea), and Munchurl Kim (Korea Advanced Institute of Science and Technology, Republic of Korea)

Gated Fields: Learning Scene Reconstruction from Gated Videos 10530
Andrea Ramazzina (Mercedes-Benz), Stefanie Walz (Mercedes-Benz, Saarland University), Pragyan Dahal (Politecnico di Milano), Mario Bijelic (Torc Robotics, Princeton University), and Felix Heide (Torc Robotics, Princeton University)

SCINeRF: Neural Radiance Fields from a Snapshot Compressive Image 10542
Yunhao Li (Zhejiang University), Xiaodong Wang (Zhejiang University), Ping Wang (Zhejiang University), Xin Yuan (Westlake University), and Peidong Liu (Westlake University)

Instance-aware Contrastive Learning for Occluded Human Mesh Reconstruction 10553
Mi-Gyeong Gwon (Konkuk University), Gi-Mun Um (Electronics and Telecommunications Research Institute), Won-Sik Cheong (Electronics and Telecommunications Research Institute), and Wonjun Kim (Konkuk University)

IBD-SLAM: Learning Image-Based Depth Fusion for Generalizable SLAM 10563
Minghao Yin (The University of Kong Kong), Shangzhe Wu (University of Oxford), and Kai Han (The University of Hong Kong)

HarmonyView: Harmonizing Consistency and Diversity in One-Image-to-3D 10574
Sangmin Woo (KAIST, Korea), Byeongjun Park (KAIST, Korea), Hyojun Go (Twelve Labs, Korea), Jin-Young Kim (Twelve Labs, Korea), and Changick Kim (KAIST, Korea)

UV-IDM: Identity-Conditioned Latent Diffusion Model for Face UV-Texture Generation 10585
Hong Li (Beihang University, China), Yutang Feng (Beihang University, China), Song Xue (Baidu VIS, China), Xuhui Liu (Beihang University, China), Bohan Zeng (Beihang University, China), Shanglin Li (Beihang University, China), Boyu Liu (Beihang University, China), Jianzhuang Liu (Shenzhen Institute of Advanced Technology, China), Shumin Han (Baidu VIS, China), and Baochang Zhang (Beihang University, China; Zhongguancun Laboratory, China; Nanchang Institute of Technology, China)

AttriHuman-3D: Editable 3D Human Avatar Generation with Attribute Decomposition and Indexing 10596
Fan Yang (Nanyang Technological University, Singapore), Tianyi Chen (South China University of Technology, China), Xiaosheng He (Nanyang Technological University, Singapore), Zhongang Cai (Nanyang Technological University, Singapore), Lei Yang (Sensetime Research, China), Si Wu (South China University of Technology, China), and Guosheng Lin (Nanyang Technological University, Singapore)

Mind The Edge: Refining Depth Edges in Sparsely-Supervised Monocular Depth Estimation 10606
Lior Talker (Samsung Israel Research Center), Aviad Cohen (Samsung Israel Research Center), Erez Yosef (Tel Aviv University), Alexandra Dana (Samsung Israel Research Center), and Michael Dinerstein (Samsung Israel Research Center)

3DiffTection: 3D Object Detection with Geometry-Aware Diffusion Features 10617
Chenfeng Xu (University of California, Berkeley, USA), Huan Ling (NVIDIA, Canada), Sanja Fidler (NVIDIA, Canada), and Or Litany (NVIDIA, Technion, Isreal)

Bayesian Diffusion Models for 3D Shape Reconstruction 10628
Haiyang Xu (University of Science and Technology of China, China), Yu Lei (Shanghai Jiao Tong University, China), Zeyuan Chen (University of California, San Diego, USA), Xiang Zhang (University of California, San Diego, USA), Yue Zhao (Tsinghua University, China), Yilin Wang (Tsinghua University, China), and Zhuowen Tu (University of California, San Diego, USA)

LaneCPP: Continuous 3D Lane Detection using Physical Priors 10639
Maximilian Pittner (Robert Bosch GmbH, Germany), Joel Janai (Robert Bosch GmbH, Germany), and Alexandru P. Condurache (Robert Bosch GmbH, Germany)

Enhancing 3D Fidelity of Text-to-3D using Cross-View Correspondences 10649
Seungwook Kim (POSTECH, Republic of Korea), Kejie Li (ByteDance, USA), Xueqing Deng (ByteDance, USA), Yichun Shi (ByteDance, USA), Minsu Cho (POSTECH, Republic of Korea), and Peng Wang (ByteDance, USA)

MonoDiff: Monocular 3D Object Detection and Pose Estimation with Diffusion Models 10659
Yasiru Ranasinghe (Johns Hopkins University, USA), Deepti Hegde (Johns Hopkins University, USA), and Vishal M. Patel (Johns Hopkins University, USA)

HiLo: Detailed and Robust 3D Clothed Human Reconstruction with High-and Low-Frequency Information of Parametric Models 10671
Yifan Yang (South China University of Technology), Dong Liu (South China University of Technology), Shuhai Zhang (South China University of Technology), Zeshuai Deng (South China University of Technology), Zixiong Huang (South China University of Technology), and Mingkui Tan (South China University of Technology)

MPOD123: One Image to 3D Content Generation Using Mask-enhanced Progressive Outline-to-Detail Optimization 10682
Jimin Xu (Zhejiang University), Tianbao Wang (Zhejiang University), Tao Jin (Zhejiang University), Shengyu Zhang (Zhejiang University), Dongjie Fu (Zhejiang University), Zhe Wang (Alibaba Group), Jiangjing Lyu (Alibaba Group), Chengfei Lv (Alibaba Group), Chaoyue Niu (Shanghai Jiaotong University), Zhou Yu (Hangzhou Dianzi University), Zhou Zhao (Zhejiang University), and Fei Wu (Zhejiang University)

GeoReF: Geometric Alignment Across Shape Variation for Category-level Object Pose Refinement 10693
Linfang Zheng (Southern University of Science and Technology, China; University of Birmingham, UK), Tze Ho Elden Tse (University of Birmingham, UK), Chen Wang (Southern University of Science and Technology, China; the University of Hong Kong, China), Yinghan Sun (Southern University of Science and Technology, China), Hua Chen (Southern University of Science and Technology, China), Ales Leonardis (University of Birmingham, UK), Wei Zhang (Southern University of Science and Technology, China), and Hyung Jin Chang (University of Birmingham, UK)

Unsupervised 3D Structure Inference from Category-Specific Image Collections 10704
Weikang Wang (University of Bonn, Germany), Dongliang Cao (University of Bonn, Germany), and Florian Bernard (University of Bonn, Germany)

Neural Parametric Gaussians for Monocular Non-Rigid Object Reconstruction 10715
Devikalyan Das (Saarland University, Germany), Christopher Wewer (Max Planck Institute for Informatics, Germany), Raza Yunus (Max Planck Institute for Informatics, Germany), Eddy Ilg (Saarland University, Germany), and Jan Eric Lenssen (Max Planck Institute for Informatics, Germany)

BiTT: Bi-directional Texture Reconstruction of Interacting Two Hands from a Single Image 10726
Minje Kim (Korea Advanced Institute of Science and Technology) and Tae-Kyun Kim (Korea Advanced Institute of Science and Technology)

DeCoTR: Enhancing Depth Completion with 2D and 3D Attentions 10736
Yunxiao Shi (Qualcomm AI Research, USA), Manish Kumar Singh (Qualcomm AI Research, USA), Hong Cai (Qualcomm AI Research, USA), and Fatih Porikli (Qualcomm AI Research, USA)

MonoNPHM: Dynamic Head Reconstruction from Monocular Videos 10747
Simon Giebenhain (Technical University of Munich), Tobias Kirschstein (Technical University of Munich), Markos Georgopoulos (Meta AI, formerly Synthesia), Martin Rünz (Synthesia), Lourdes Agapito (University College London), and Matthias Nießner (Technical University of Munich)

FakeInversion: Learning to Detect Images from Unseen Text-to-Image Models by Inverting Stable Diffusion 10759
George Cazenavette (MIT), Avneesh Sud (Google Research), Thomas Leung (Google Research), and Ben Usman (Google Research)

Forgery-aware Adaptive Transformer for Generalizable Synthetic Image Detection 10770
Huan Liu (Beijing Jiaotong University, China), Zichang Tan (Baidu, China), Chuangchuang Tan (Beijing Jiaotong University, China), Yunchao Wei (Beijing Jiaotong University, China), Jingdong Wang (Baidu, China), and Yao Zhao (Beijing Jiaotong University, China)

Towards Modern Image Manipulation Localization: A Large-Scale Dataset and Novel Methods 10781
Chenfan Qu (South China University of Technology), Yiwu Zhong (University of Wisconsin), Chongyu Liu (South China University of Technology), Guitao Xu (South China University of Technology), Dezhi Peng (South China University of Technology), Fengjun Guo (INTSIG), and Lianwen Jin (South China University of Technology)

ManiFPT: Defining and Analyzing Fingerprints of Generative Models 10791
Hae Jin Song (University of Southern California), Mahyar Khayatkhoei (University of Southern California), and Wael AbdAlmageed (Clemson University)

ProMark: Proactive Diffusion Watermarking for Causal Attribution 10802
Vishal Asnani (Adobe Research, Michigan State University), John Collomosse (Adobe Research, University of Surrey), Tu Bui (University of Surrey), Xiaoming Liu (Michigan State University), and Shruti Agarwal (Adobe Research)

CGI-DM: Digital Copyright Authentication for Diffusion Models via Contrasting Gradient Inversion 10812
Xiaoyu Wu (Shanghai Jiao Tong University), Yang Hua (Queen's University Belfast), Chumeng Liang (University of Southern California), Jiaru Zhang (Shanghai Jiao Tong University), Hao Wang (Louisiana State University), Tao Song (Shanghai Jiao Tong University), and Haibing Guan (Shanghai Jiao Tong University)

SCoFT: Self-Contrastive Fine-Tuning for Equitable Image Generation 10822
Zhixuan Liu (Carnegie Mellon University), Peter Schaldenbrand (Carnegie Mellon Univeristy), Beverley-Claire Okogwu (Carnegie Mellon University), Wenxuan Peng (Nanyang Technological University), Youngsik Yun (Dongguk University), Andrew Hundt (Carnegie Mellon University), Jihie Kim (Dongguk University), and Jean Oh (Carnegie Mellon University)

Would Deep Generative Models Amplify Bias in Future Models? 10833
Tianwei Chen (Osaka University), Yusuke Hirota (Osaka University), Mayu Otani (CyberAgent Inc.), Noa Garcia (Osaka University), and Yuta Nakashima (Osaka University)

Training Diffusion Models Towards Diverse Image Generation with Reinforcement Learning 10844
Zichen Miao (Purdue University), Jiang Wang (Microsoft Corporation), Ze Wang (Purdue University), Zhengyuan Yang (Microsoft Corporation), Lijuan Wang (Microsoft Corporation), Qiang Qiu (Purdue University), and Zicheng Liu (Advanced Micro Devices, Inc.)

Consistency and Uncertainty: Identifying Unreliable Responses From Black-Box Vision-Language Models for Selective Visual Question Answering 10854
Zaid Khan (Northeastern University) and Yun Fu (Northeastern University)

Visual Objectification in Films: Towards a New AI Task for Video Interpretation 10864
Julie Tores (Université Côte d'Azur, France), Lucile Sassatelli (Université Côte d'Azur, France), Hui-Yin Wu (Université Côte d'Azur, Inria, France), Clement Bergman (Université Côte d'Azur, Inria, France), Léa Andolfi (Sorbonne Université, France), Victor Ecrement (Sorbonne Université, France), Frédéric Precioso (Université Côte d'Azur, Inria, France), Thierry Devars (Sorbonne Université, France), Magali Guaresi (Université Côte d'Azur, France), Virginie Julliard (Sorbonne Université, France), and Sarah Lecossais (Université Sorbonne Paris Nord, France)

ToonerGAN: Reinforcing GANs for Obfuscating Automated Facial Indexing 10875
Kartik Thakral (IIT Jodhpur), Shashikant Prasad (IIT Jodhpur), Stuti Aswani (IIT Jodhpur), Mayank Vatsa (IIT Jodhpur), and Richa Singh (IIT Jodhpur)

MCPNet: An Interpretable Classifier via Multi-Level Concept Prototypes 10885
Bor-Shiun Wang (National Yang Ming Chiao Tung University), Chien-Yi Wang (NVIDIA Research), and Wei-Chen Chiu (National Yang Ming Chiao Tung University)

Visual Concept Connectome (VCC): Open World Concept Discovery and their Interlayer Connections in Deep Models 10895
Matthew Kowal (York University, Canada), Richard P. Wildes (York University, Canada), and Konstantinos G. Derpanis (York University, Canada)

Discover and Mitigate Multiple Biased Subgroups in Image Classifiers 10906
Zeliang Zhang (University of Rochester, USA), Mingqian Feng (University of Rochester, USA), Zhiheng Li (University of Rochester, USA), and Chenliang Xu (University of Rochester, USA)

CORES: Convolutional Response-based Score for Out-of-distribution Detection 10916
Keke Tang (Guangzhou University), Chao Hou (Guangzhou University), Weilong Peng (Guangzhou University), Runnan Chen (University of Hong Kong), Peican Zhu (Northwestern Polytechnical University), Wenping Wang (Texas A&M University), and Zhihong Tian (Guangzhou University)

Token Transformation Matters: Towards Faithful Post-hoc Explanation for Vision Transformer.. 10926
Junyi Wu (Illinois Institute of Technology, USA), Bin Duan (Illinois Institute of Technology, USA), Weitai Kang (Illinois Institute of Technology, USA), Hao Tang (Carnegie Mellon University, USA), and Yan Yan (Illinois Institute of Technology, USA)

On the Faithfulness of Vision Transformer Explanations 10936
Junyi Wu (Illinois Institute of Technology, USA), Weitai Kang (Illinois Institute of Technology, USA), Hao Tang (Carnegie Mellon University, USA), Yuan Hong (University of Connecticut, USA), and Yan Yan (Illinois Institute of Technology, USA)

Understanding Video Transformers via Universal Concept Discovery 10946
Matthew Kowal (York University, Canada), Achal Dave (Toyota Research Institute, USA), Rares Ambrus (Toyota Research Institute, USA), Adrien Gaidon (Toyota Research Institute, USA), Konstantinos G. Derpanis (York University, Canada), and Pavel Tokmakov (Toyota Research Institute, USA)

Explaining the Implicit Neural Canvas: Connecting Pixels to Neurons by Tracing their Contributions 10957
Namitha Padmanabhan (University of Maryland, College Park, USA), Matthew Gwilliam (University of Maryland, College Park, USA), Pulkit Kumar (University of Maryland, College Park, USA), Shishira R Maiya (University of Maryland, College Park, USA), Max Ehrlich (University of Maryland, College Park, USA), and Abhinav Shrivastava (University of Maryland, College Park, USA)

WWW: A Unified Framework for Explaining What, Where and Why of Neural Networks by Interpretation of Neuron Concepts 10968
Yong Hyun Ahn (Kyung Hee University, Republic of Korea), Hyeon Bae Kim (Kyung Hee University, Republic of Korea), and Seong Tae Kim (Kyung Hee University, Republic of Korea)

HDQMF: Holographic Feature Decomposition Using Quantum Algorithms 10978
Prathyush Prasanth Poduval (University of California, Irvine), Zhuowen Zou (University of California, Irvine), and Mohsen Imani (University of California, Irvine)

SLICE: Stabilized LIME for Consistent Explanations for Image Classification 10988
Revoti Prasad Bora (Norwegian University of Science and Technology), Philipp Terhörst (University of Paderborn), Raymond Veldhuis (Norwegian University of Science and Technology), Raghavendra Ramachandra (Norwegian University of Science and Technology), and Kiran Raja (Norwegian University of Science and Technology)

What Sketch Explainability Really Means for Downstream Tasks? 10997
Hmrishav Bandyopadhyay (University of Surrey), Pinaki Nath Chowdhury (University of Surrey), Ayan Kumar Bhunia (University of Surrey), Aneeshan Sain (University of Surrey), Tao Xiang (University of Surrey, iFlyTek-Surrey Joint Research Centre on Artificial Intelligence), and Yi-Zhe Song (University of Surrey, iFlyTek-Surrey Joint Research Centre on Artificial Intelligence)

Structured Gradient-based Interpretations via Norm-Regularized Adversarial Training 11009
Shizhan Gong (The Chinese University of Hong Kong, China), Qi Dou (The Chinese University of Hong Kong, China), and Farzan Farnia (The Chinese University of Hong Kong, China)

Learning Triangular Distribution in Visual World 11019
Ping Chen (MicroBT Inc.), Xingpeng Zhang (Southwest Petroleum University, Chengdu, China), Chengtao Zhou (MicroBT Inc.), Dichao Fan (MicroBT Inc.), Peng Tu (RuqiMobility Inc.), Le Zhang (MicroBT Inc.), and Yanlin Qian (MicroBT Inc.)

Incremental Residual Concept Bottleneck Models 11030
Chenming Shang (Tsinghua University), Shiji Zhou (Tsinghua University), Hengyuan Zhang (Tsinghua University), Xinzhe Ni (Tsinghua University), Yujiu Yang (Tsinghua University), and Yuwang Wang (Tsinghua University)

Uncertainty Visualization via Low-Dimensional Posterior Projections 11041
Omer Yair (Technion, Israel), Elias Nehme (Technion, Israel), and Tomer Michaeli (Technion, Israel)

Epistemic Uncertainty Quantification For Pre-Trained Neural Networks 11052
Hanjing Wang (Rensselaer Polytechnic Institute, USA) and Qiang Ji (Rensselaer Polytechnic Institute, USA)

Interpretable Measures of Conceptual Similarity by Complexity-Constrained Descriptive Auto-Encoding 11062
Alessandro Achille (AWS AI Labs), Greg Ver Steeg (AWS AI Labs), Tian Yu Liu (AWS AI Labs), Matthew Trager (AWS AI Labs), Carson Klingenberg (AWS AI Labs), and Stefano Soatto (AWS AI Labs)

CAPE: CAM as a Probabilistic Ensemble for Enhanced DNN Interpretation 11072
Townim Faisal Chowdhury (Australian Institute for Machine Learning, University of Adelaide), Kewen Liao (Australian Catholic University), Vu Minh Hieu Phan (Australian Institute for Machine Learning, University of Adelaide), Minh-Son To (Flinders University), Yutong Xie (Australian Institute for Machine Learning, University of Adelaide), Kevin Hung (SA Pathology, Central Adelaide Local Health Network), David Ross (SA Pathology, Central Adelaide Local Health Network), Anton van den Hengel (Australian Institute for Machine Learning, University of Adelaide), Johan W. Verjans (Australian Institute for Machine Learning, University of Adelaide), and Zhibin Liao (Australian Institute for Machine Learning, University of Adelaide)

Discovering and Mitigating Visual Biases through Keyword Explanation 11082
Younghyun Kim (KAIST), Sangwoo Mo (University of Michigan), Minkyu Kim (KRAFTON), Kyungmin Lee (KAIST), Jaeho Lee (POSTECH), and Jinwoo Shin (KAIST)

DiG-IN: Diffusion Guidance for Investigating Networks - Uncovering Classifier Differences, Neuron Visualisations, and Visual Counterfactual Explanations 11093
Maximilian Augustin (University of Tübingen, Germany), Yannic Neuhaus (University of Tübingen, Germany), and Matthias Hein (University of Tübingen, Germany)

Cross-Dimension Affinity Distillation for 3D EM Neuron Segmentation 11104
Xiaoyu Liu (MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition, University of Science and Technology of China), Miaomiao Cai (MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition, University of Science and Technology of China), Yinda Chen (MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition, University of Science and Technology of China), Yueyi Zhang (MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition, University of Science and Technology of China), Te Shi (Institute of Artificial Intelligence, Hefei Comprehensive National Science Center), Ruobing Zhang (Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences), Xuejin Chen (MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition, University of Science and Technology of China), and Zhiwei Xiong (MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition, University of Science and Technology of China)

Continual Self-supervised Learning: Towards Universal Multi-modal Medical Data Representation Learning 11114
Yiwen Ye (Northwestern Polytechnical University, China), Yutong Xie (University of Adelaide, Australia), Jianpeng Zhang (Zhejiang University, China), Ziyang Chen (Northwestern Polytechnical University, China), Qi Wu (University of Adelaide, Australia), and Yong Xia (Northwestern Polytechnical University, China; Research & Development Institute of Northwestern Polytechnical University in Shenzhen, China; Ningbo Institute of Northwestern Polytechnical University, China)

A Unified Framework for Microscopy Defocus Deblur with Multi-Pyramid Transformer and Contrastive Learning 11125
Yuelin Zhang (The Chinese University of Hong Kong, Hong Kong SAR), Pengyu Zheng (The Chinese University of Hong Kong, Hong Kong SAR), Wanquan Yan (The Chinese University of Hong Kong, Hong Kong SAR), Chengyu Fang (Tsinghua University, China), and Shing Shin Cheng (The Chinese University of Hong Kong, Hong Kong SAR)

CARZero: Cross-Attention Alignment for Radiology Zero-Shot Classification 11137
Haoran Lai (University of Science and Technology of China, China), Qingsong Yao (Chinese Academy of Science, China), Zihang Jiang (University of Science and Technology of China, China), Rongsheng Wang (University of Science and Technology of China, China), Zhiyang He (iFlytek Co.Ltd, China), Xiaodong Tao (iFlytek Co.Ltd, China), and S. Kevin Zhou (University of Science and Technology of China, China)

Towards Generalizable Tumor Synthesis 11147
Qi Chen (University of Science and Technology of China), Xiaoxi Chen (Shanghai Jiao Tong University), Haorui Song (Johns Hopkins University), Zhiwei Xiong (University of Science and Technology of China), Alan Yuille (Johns Hopkins University), Chen Wei (Johns Hopkins University), and Zongwei Zhou (Johns Hopkins University)

Tyche: Stochastic In-Context Learning for Medical Image Segmentation 11159
Marianne Rakic (Massachusetts Institute of Technology, USA), Hallee E. Wong (Massachusetts Institute of Technology, USA), Jose Javier Gonzalez Ortiz (MosaicML DataBricks & Massachusetts Institute of Technology, USA), Beth A. Cimini (Broad Institute of MIT and Harvard, USA), John V. Guttag (Massachusetts Institute of Technology, USA), and Adrian V. Dalca (Massachusetts Institute of Technology & Harvard Medical School, USA)

Structure-Aware Sparse-View X-ray 3D Reconstruction 11174
Yuanhao Cai (Johns Hopkins University, USA), Jiahao Wang (Johns Hopkins University, USA), Alan Yuille (Johns Hopkins University, USA), Zongwei Zhou (Johns Hopkins University, USA), and Angtian Wang (Johns Hopkins University, USA)

Each Test Image Deserves A Specific Prompt: Continual Test-Time Adaptation for 2D Medical Image Segmentation 11184
Ziyang Chen (Northwestern Polytechnical University, China), Yongsheng Pan (Northwestern Polytechnical University, China), Yiwen Ye (Northwestern Polytechnical University, China), Mengkang Lu (Northwestern Polytechnical University, China), and Yong Xia (Northwestern Polytechnical University, China)

Training Like a Medical Resident: Context-Prior Learning Toward Universal Medical Image Segmentation 11194
Yunhe Gao (Rutgers University)

C^2RV: Cross-Regional and Cross-View Learning for Sparse-View CBCT Reconstruction 11205
Yiqun Lin (The Hong Kong University of Science and Technology), Jiewen Yang (The Hong Kong University of Science and Technology), Hualiang Wang (The Hong Kong University of Science and Technology), Xinpeng Ding (The Hong Kong University of Science and Technology), Wei Zhao (Beihang University), and Xiaomeng Li (The Hong Kong University of Science and Technology)

Modality-Agnostic Structural Image Representation Learning for Deformable Multi-Modality Medical Image Registration 11215
Tony C. W. Mok (DAMO Academy, Alibaba Group), Zi Li (DAMO Academy, Alibaba Group), Yunhao Bai (East China Normal University), Jianpeng Zhang (DAMO Academy, Alibaba Group), Wei Liu (DAMO Academy, Alibaba Group), Yan-Jie Zhou (DAMO Academy, Alibaba Group), Ke Yan (DAMO Academy, Alibaba Group), Dakai Jin (DAMO Academy, Alibaba Group), Yu Shi (China Medical University Shenyang), Xiaoli Yin (China Medical University Shenyang), Le Lu (DAMO Academy, Alibaba Group), and Ling Zhang (DAMO Academy, Alibaba Group)

SI-MIL: Taming Deep MIL for Self-Interpretability in Gigapixel Histopathology 11226
Saarthak Kapse (Stony Brook University, USA), Pushpak Pati (Independent Researcher), Srijan Das (UNC Charlotte, USA), Jingwei Zhang (Stony Brook University, USA), Chao Chen (Stony Brook University, USA), Maria Vakalopoulou (CentraleSupelec, University of Paris-Saclay, France), Joel Saltz (Stony Brook University, USA), Dimitris Samaras (Stony Brook University, USA), Rajarsi R. Gupta (Stony Brook University, USA), and Prateek Prasanna (Stony Brook University, USA)

Bootstrapping Chest CT Image Understanding by Distilling Knowledge from X-ray Expert Models 11238
Weiwei Cao (DAMO Academy, Alibaba Group; University of Science and Technology of China, China), Jianpeng Zhang (DAMO Academy, Alibaba Group; Zhejiang University, China), Yingda Xia (DAMO Academy, Alibaba Group), Tony C. W. Mok (DAMO Academy, Alibaba Group), Zi Li (DAMO Academy, Alibaba Group), Xianghua Ye (Zhejiang University, China), Le Lu (DAMO Academy, Alibaba Group), Jian Zheng (University of Science and Technology of China, China), Yuxing Tang (DAMO Academy, Alibaba Group), and Ling Zhang (DAMO Academy, Alibaba Group)

ViLa-MIL: Dual-scale Vision-Language Multiple Instance Learning for Whole Slide Image Classification 11248
Jiangbo Shi (Xi'an Jiaotong University, China), Chen Li (Xi'an Jiaotong University, China), Tieliang Gong (Xi'an Jiaotong University, China), Yefeng Zheng (Tencent YouTu Lab, China), and Huazhu Fu (Agency for Science, Technology and Research, Singapore)

Virtual Immunohistochemistry Staining for Histological Images Assisted by Weakly-supervised Learning 11259
Jiahan Li (Harbin Institute of Technology), Jiuyang Dong (Harbin Institute of Technology), Shenjin Huang (Harbin Institute of Technology), Xi Li (Peking University Shenzhen Hospital), Junjun Jiang (Harbin Institute of Technology), Xiaopeng Fan (Harbin Institute of Technology), and Yongbing Zhang (Harbin Institute of Technology, Shenzhen)

Representing Part-Whole Hierarchies in Foundation Models by Learning Localizability, Composability, and Decomposability from Anatomy via Self Supervision 11269
Mohammad Reza Hosseinzadeh Taher (Arizona State University, USA), Michael B. Gotway (Mayo Clinic, USA), and Jianming Liang (Arizona State University, USA)

XFibrosis: Explicit Vessel-Fiber Modeling for Fibrosis Staging from Liver Pathology Images..... 11282
Chong Yin (Hong Kong Baptist University, Hong Kong), Siqi Liu (Shenzhen Research Institute of Big Data, Chinese University of Hong Kong, Shenzhen), Fei Lyu (Hong Kong Baptist University, Hong Kong), Jiahao Lu (University of Copenhagen, Denmark), Sune Darkner (University of Copenhagen, Denmark), Vincent Wai-Sun Wong (Chinese University of Hong Kong, Hong Kong), and Pong C. Yuen (Hong Kong Baptist University, Hong Kong)

Prompting Vision Foundation Models for Pathology Image Analysis 11292
Chong Yin (Hong Kong Baptist University, Hong Kong), Siqi Liu (Shenzhen Research Institute of Big Data, China), Kaiyang Zhou (Hong Kong Baptist University, Hong Kong), Vincent Wai-Sun Wong (Chinese University of Hong Kong, Hong Kong), and Pong C. Yuen (Hong Kong Baptist University, Hong Kong)

One-Prompt to Segment All Medical Images 11302
Junde Wu (The University of Oxford) and Min Xu (Carnegie Mellon University)

Learning Large-Factor EM Image Super-Resolution with Generative Priors 11313
Jiateng Shou (University of Science and Technology of China, China), Zeyu Xiao (University of Science and Technology of China, China), Shiyu Deng (University of Science and Technology of China, China), Wei Huang (University of Science and Technology of China, China), Peiyao Shi (Suzhou Institute of Biomedical Engineering and Technology, China), Ruobing Zhang (Suzhou Institute of Biomedical Engineering and Technology, China), Zhiwei Xiong (University of Science and Technology of China, China), and Feng Wu (University of Science and Technology of China, China)

Dynamic Graph Representation with Knowledge-aware Attention for Histopathology Whole Slide Image Analysis 11323
Jiawen Li (Tsinghua University, China), Yuxuan Chen (Tsinghua University, China), Hongbo Chu (Tsinghua University, China), Qiehe Sun (Tsinghua University, China), Tian Guan (Tsinghua University, China), Anjia Han (The First Affiliated Hospital of Sun Yat-sen University, China), and Yonghong He (Tsinghua University, China)

MindBridge: A Cross-Subject Brain Decoding Framework 11333
Shizun Wang (National University of Singapore), Songhua Liu (National University of Singapore), Zhenxiong Tan (National University of Singapore), and Xinchao Wang (National University of Singapore)

Feature Re-Embedding: Towards Foundation Model-Level Performance in Computational Pathology 11343
Wenhao Tang (Chonqing University, China), Fengtao Zhou (The Hong Kong University of Science and Technology, China), Sheng Huang (Chonqing University, China), Xiang Zhu (Chonqing University, China), Yi Zhang (Chonqing University, China), and Bo Liu (Walmart Global Tech, USA)

Data-Efficient Unsupervised Interpolation Without Any Intermediate Frame for 4D Medical Images 11353
JungEun Kim (Korea Advanced Institute of Science and Technology (KAIST), Republic of Korea), Hangyul Yoon (Korea Advanced Institute of Science and Technology (KAIST), Republic of Korea), Geondo Park (Korea Advanced Institute of Science and Technology (KAIST), Republic of Korea), Kyungsu Kim (Massachusetts General Hospital and Harvard Medical School, USA), and Eunho Yang (Korea Advanced Institute of Science and Technology (KAIST), Republic of Korea)

Rethinking Diffusion Model for Multi-Contrast MRI Super-Resolution 11365
Guangyuan Li (Zhejiang University, China), Chen Rao (Zhejiang University, China), Juncheng Mo (Zhejiang University, China), Zhanjie Zhang (Zhejiang University, China), Wei Xing (Zhejiang University, China), and Lei Zhao (Zhejiang University, China)

Adapting Visual-Language Models for Generalizable Anomaly Detection in Medical Images 11375
Chaoqin Huang (Shanghai Jiao Tong University; National University of Singapore; Shanghai Artificial Intelligence Laboratory), Aofan Jiang (Shanghai Jiao Tong University; Shanghai Artificial Intelligence Laboratory), Jinghao Feng (Shanghai Jiao Tong University; Shanghai Artificial Intelligence Laboratory), Ya Zhang (Shanghai Jiao Tong University; Shanghai Artificial Intelligence Laboratory), Xinchao Wang (National University of Singapore), and Yanfeng Wang (Shanghai Jiao Tong University; Shanghai Artificial Intelligence Laboratory)

ZePT: Zero-Shot Pan-Tumor Segmentation via Query-Disentangling and Self-Prompting 11386
Yankai Jiang (Shanghai AI Laboratory, China), Zhongzhen Huang (Shanghai AI Laboratory, China; Shanghai Jiao Tong University, China), Rongzhao Zhang (Shanghai AI Laboratory, China), Xiaofan Zhang (Shanghai AI Laboratory, China; Shanghai Jiao Tong University, China), and Shaoting Zhang (Shanghai AI Laboratory, China; SenseTime Research, China)

Generalizable Whole Slide Image Classification with Fine-Grained Visual-Semantic Interaction 11398
Hao Li (Xiamen University, China), Ying Chen (Xiamen University, China), Yifei Chen (Huawei, China), Rongshan Yu (Xiamen University, China), Wenxian Yang (Aginome Scientific, China), Liansheng Wang (Xiamen University, China), Bowen Ding (Shanghai Jiao Tong University School of Medicine, China), and Yuchen Han (Shanghai Jiao Tong University School of Medicine, China)

Incremental Nuclei Segmentation from Histopathological Images via Future-class Awareness and Compatibility-inspired Distillation 11408
Huyong Wang (Shenzhen University), Huisi Wu (Shenzhen University), and Jing Qin (The Hong Kong Polytechnic University)

PH-Net: Semi-Supervised Breast Lesion Segmentation via Patch-wise Hardness 11418
Siyao Jiang (Shenzhen University, China), Huisi Wu (Shenzhen University, China), Junyang Chen (Shenzhen University, China), Qin Zhang (Shenzhen University, China), and Jing Qin (The Hong Kong Polytechnic University, Hong Kong)

ToNNO: Tomographic Reconstruction of a Neural Network's Output for Weakly Supervised Segmentation of 3D Medical Images 11428
Marius Schmidt-Mengin (TheraPanacea, France), Alexis Benichoux (TheraPanacea, France), Shibeshih Belachew (TheraPanacea, France), Nikos Komodakis (University of Crete, Greece), and Nikos Paragios (TheraPanacea, France)

Think Twice Before Selection: Federated Evidential Active Learning for Medical Image Analysis with Domain Shifts 11439
Jiayi Chen (Northwestern Polytechnical University, China), Benteng Ma (Hong Kong University of Science and Technology, Hong Kong SAR, China), Hengfei Cui (Northwestern Polytechnical University, China), and Yong Xia (Northwestern Polytechnical University, China; Research & Development Institute of Northwestern Polytechnical University in Shenzhen, China)

CPLIP: Zero-Shot Learning for Histopathology with Comprehensive Vision-Language Alignment 11450
Sajid Javed (Khalifa University of Science and Technology, UAE), Arif Mahmood (Information Technology University of the Punjab, Pakistan), Iyyakutti Iyappan Ganapathi (Khalifa University of Science and Technology, UAE), Fayaz Ali Dharejo (Khalifa University of Science and Technology, UAE), Naoufel Werghi (Khalifa University of Science and Technology, UAE), and Mohammed Bennamoun (The University of Western Australia, Australia)

MicroDiffusion: Implicit Representation-Guided Diffusion for 3D Reconstruction from Limited 2D Microscopy Projections 11460
Mude Hui (University of California, Santa Cruz), Zihao Wei (University of Michigan, Ann Arbor), Hongru Zhu (Johns Hopkins University), Fei Xia (Ecole Normale Supérieure de Paris), and Yuyin Zhou (University of California, Santa Cruz)

Diversified and Personalized Multi-rater Medical Image Segmentation 11470
Yicheng Wu (Monash University), Xiangde Luo (University of Electronic Science and Technology of China), Zhe Xu (The Chinese University of Hong Kong), Xiaoqing Guo (University of Oxford), Lie Ju (Monash University), Zongyuan Ge (Monash University), Wenjun Liao (University of Electronic Science and Technology of China), and Jianfei Cai (Monash University)

Modality-agnostic Domain Generalizable Medical Image Segmentation by Multi-Frequency in Multi-Scale Attention 11480
Ju-Hyeon Nam (Inha University), Nur Suriza Syazwany (Inha University), Su Jung Kim (Inha University), and Sang-Chul Lee (Inha University)

Decomposing Disease Descriptions for Enhanced Pathology Detection: A Multi-Aspect Vision-Language Pre-training Framework 11492
Vu Minh Hieu Phan (The University of Adelaide, Australia), Yutong Xie (The University of Adelaide, Australia), Yuankai Qi (Macquarie University, Australia), Lingqiao Liu (The University of Adelaide, Australia), Liyang Liu (The University of Adelaide, Australia), Bowen Zhang (The University of Adelaide, Australia), Zhibin Liao (The University of Adelaide, Australia), Qi Wu (The University of Adelaide, Australia), Minh-Son To (Flinders University, Australia), and Johan W. Verjans (The University of Adelaide, Australia)

MedM2G: Unifying Medical Multi-Modal Generation via Cross-Guided Diffusion with Visual Invariant 11502
Chenlu Zhan (Zhejiang University), Yu Lin (Zhejiang University), Gaoang Wang (Zhejiang University), Hongwei Wang (Zhejiang University), and Jian Wu (Zhejiang University)

H-ViT: A Hierarchical Vision Transformer for Deformable Image Registration 11513
Morteza Ghahremani (Technical University of Munich, Germany; Munich Center for Machine Learning, Germany), Mohammad Khateri (University of Eastern Finland, Finland), Bailiang Jian (Technical University of Munich, Germany; Munich Center for Machine Learning, Germany), Benedikt Wiestler (Technical University of Munich, Germany), Ehsan Adeli (Stanford University, USA), and Christian Wachinger (Technical University of Munich, Germany; Munich Center for Machine Learning, Germany)

Seeing Unseen: Discover Novel Biomedical Concepts via Geometry-Constrained Probabilistic Modeling 11524
Jianan Fan (University of Sydney), Dongnan Liu (University of Sydney), Hang Chang (Lawrence Berkeley National Laboratory), Heng Huang (University of Maryland at College Park), Mei Chen (Microsoft), and Weidong Cai (University of Sydney)

Fully Convolutional Slice-to-Volume Reconstruction for Single-Stack MRI 11535
Sean I. Young (Harvard Medical School, USA), Yael Balbastre (Harvard Medical School, USA), Bruce Fischl (Harvard Medical School, USA), Polina Golland (MIT, USA), and Juan Eugenio Iglesias (Harvard Medical School, USA)

IIRP-Net: Iterative Inference Residual Pyramid Network for Enhanced Image Registration 11546
Tai Ma (East China Normal University), Suwei Zhang (East China Normal University), Jiafeng Li (East China Normal University), and Ying Wen (East China Normal University)

ChAda-ViT : Channel Adaptive Attention for Joint Representation Learning of Heterogeneous Microscopy Images 11556
Nicolas Bourriez (Ecole Normale Supérieure PSL, France), Ihab Bendidi (Ecole Normal Supérieure PSL, France), Ethan Cohen (Ecole Normale Supérieure PSL, France), Gabriel Watkinson (Ecole Normale Supérieure PSL, France), Maxime Sanchez (Ecole Normale Supérieure PSL, France), Guillaume Bollot (Synsight, France), and Auguste Genovesio (Ecole Normale Supérieure PSL, France)

Morphological Prototyping for Unsupervised Slide Representation Learning in Computational Pathology 11566
Andrew H. Song (Mass General Brigham, USA), Richard J. Chen (Mass General Brigham, USA), Tong Ding (Mass General Brigham, USA), Drew F.K. Williamson (Mass General Brigham, USA), Guillaume Jaume (Mass General Brigham, USA), and Faisal Mahmood (Mass General Brigham, USA)

Modeling Dense Multimodal Interactions Between Biological Pathways and Histology for Survival Prediction 11579
Guillaume Jaume (Harvard Medical School, USA), Anurag Vaidya (Harvard Medical School, USA), Richard J. Chen (Harvard Medical School, USA), Drew F.K. Williamson (Mass General Brigham, USA), Paul Pu Liang (Carnegie Mellon University, USA), and Faisal Mahmood (Harvard Medical School, USA)

Accurate Spatial Gene Expression Prediction by Integrating Multi-Resolution Features 11591
Youngmin Chung (SungKyunKwan University, South Korea), Ji Hun Ha (SungKyunKwan University, South Korea), Kyeong Chan Im (SungKyunKwan University, South Korea), and Joo Sang Lee (SungKyunKwan University, South Korea)

Teeth-SEG: An Efficient Instance Segmentation Framework for Orthodontic Treatment based on Multi-Scale Aggregation and Anthropic Prior Knowledge 11601
Bo Zou (Tsinghua University, China), Shaofeng Wang (Capital Medical Universty, China), Hao Liu (Tsinghua University, China), Gaoyue Sun (Imperial College London, England), Yajie Wang (Tsinghua University, China), FeiFei Zuo (LargeV .Inc, China), Chengbin Quan (Tsinghua University, China), and Youjian Zhao (Tsinghua University, China)

Low-Rank Knowledge Decomposition for Medical Foundation Models 11611
Yuhang Zhou (Shanghai Jiao Tong University, China; Shanghai Artificial Intelligence Laboratory, China), Haolin Li (Fudan University, China; Shanghai Artificial Intelligence Laboratory, China), Siyuan Du (Fudan University, China; Shanghai Artificial Intelligence Laboratory, China), Jiangchao Yao (Shanghai Jiao Tong University, China; Shanghai Artificial Intelligence Laboratory, China), Ya Zhang (Shanghai Jiao Tong University, China; Shanghai Artificial Intelligence Laboratory, China), and Yanfeng Wang (Shanghai Jiao Tong University, China; Shanghai Artificial Intelligence Laboratory, China)

M3-UDA: A New Benchmark for Unsupervised Domain Adaptive Fetal Cardiac Structure Detection... 11621
Bin Pu (The Hong Kong University of Science and Technology), Liwen Wang (Anhui University), Jiewen Yang (The Hong Kong University of Science and Technology), Guannan He (Sichuan Provincial Maternity and Child Health Care Hospital), Xingbo Dong (Anhui University), Shengli Li (Shenzhen Maternity and Child Healthcare Hospital), Ying Tan (Shenzhen Maternity and Child Healthcare Hospital), Ming Chen (Harbin Red Cross Central Hospital), Zhe Jin (Anhui University), Kenli Li (Hunan University), and Xiaomeng Li (The Hong Kong University of Science and Technology)

CycleINR: Cycle Implicit Neural Representation for Arbitrary-Scale Volumetric Super-Resolution of Medical Data 11631
Wei Fang (Alibaba Group, China), Yuxing Tang (Alibaba Group, USA), Heng Guo (Alibaba Group, China), Mingze Yuan (Alibaba Group, China), Tony C. W. Mok (Alibaba Group, China), Ke Yan (Alibaba Group, China), Jiawen Yao (Alibaba Group, China), Xin Chen (Guangdong Provincial People's Hospital, China), Zaiyi Liu (Guangdong Provincial People's Hospital, China), Le Lu (Alibaba Group, USA), Ling Zhang (Alibaba Group, USA), and Minfeng Xu (Alibaba Group, China)

Constructing and Exploring Intermediate Domains in Mixed Domain Semi-supervised Medical Image Segmentation 11642
Qinghe Ma (Nanjing University, China), Jian Zhang (Nanjing University, China), Lei Qi (Southeast University, China), Qian Yu (Shandong Women's University, China), Yinghuan Shi (Nanjing University, China), and Yang Gao (Nanjing University, China)

PairAug: What Can Augmented Image-Text Pairs Do for Radiology? 11652
Yutong Xie (The University of Adelaide, Australia), Qi Chen (The University of Adelaide, Australia), Sinuo Wang (The University of Adelaide, Australia), Minh-Son To (South Australia Medical Imaging, Australia), Iris Lee (South Australia Medical Imaging, Australia), Ee Win Khoo (South Australia Medical Imaging, Australia), Kerolos Hendy (South Australia Medical Imaging, Australia), Daniel Koh (South Australia Medical Imaging, Australia), Yong Xia (Northwestern Polytechnical University, China), and Qi Wu (The University of Adelaide, Australia)

Intraoperative 2D/3D Image Registration via Differentiable X-ray Rendering 11662
Vivek Gopalakrishnan (Massachusetts Institute of Technology, USA), Neel Dey (Massachusetts Institute of Technology, USA), and Polina Golland (Massachusetts Institute of Technology, USA)

Mudslide: A Universal Nuclear Instance Segmentation Method 11673
Jun Wang (Peking University, China)

Rotation-Agnostic Image Representation Learning for Digital Pathology 11683
Saghir Alfasly (KIMIA Lab, Mayo Clinic, Rochester, MN, USA), Abubakr Shafique (KIMIA Lab, Mayo Clinic, Rochester, MN, USA), Peyman Nejat (KIMIA Lab, Mayo Clinic, Rochester, MN, USA), Jibran Khan (KIMIA Lab, Mayo Clinic, Rochester, MN, USA), Areej Alsaafin (KIMIA Lab, Mayo Clinic, Rochester, MN, USA), Ghazal Alabtah (KIMIA Lab, Mayo Clinic, Rochester, MN, USA), and H.R. Tizhoosh (KIMIA Lab, Mayo Clinic, Rochester, MN, USA)

Tumor Micro-environment Interactions Guided Graph Learning for Survival Analysis of Human Cancers from Whole-slide Pathological Images 11694
Wei Shao (Nanjing University of Aeronautics and Astronautics, China; Key Laboratory of Brain-Machine Intelligence Technology, China; MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, China), YangYang Shi (Nanjing University of Aeronautics and Astronautics, China; Key Laboratory of Brain-Machine Intelligence Technology, China; MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, China), Daoqiang Zhang (Nanjing University of Aeronautics and Astronautics, China; Key Laboratory of Brain-Machine Intelligence Technology, China; MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, China), JunJie Zhou (Nanjing University of Aeronautics and Astronautics, China; Key Laboratory of Brain-Machine Intelligence Technology, China; MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, China), and Peng Wan (Nanjing University of Aeronautics and Astronautics, China; Key Laboratory of Brain-Machine Intelligence Technology, China; MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, China)

MLIP: Enhancing Medical Visual Representation with Divergence Encoder and Knowledge-guided Contrastive Learning 11704
Zhe Li (Huazhong University of Science and Technology), Laurence T. Yang (Huazhong University of Science and Technology; Zhengzhou University), Bocheng Ren (Huazhong University of Science and Technology), Xin Nie (Huazhong University of Science and Technology), Zhangyang Gao (Westlake University), Cheng Tan (Westlake University), and Stan Z. Li (Westlake University)

FocusMAE: Gallbladder Cancer Detection from Ultrasound Videos with Focused Masked Autoencoders 11715
Soumen Basu (Indian Institute of Technology Delhi, India), Mayuna Gupta (Indian Institute of Technology Delhi, India), Chetan Madan (Indian Institute of Technology Delhi, India), Pankaj Gupta (Postgraduate Institute of Medical Education and Research, Chandigarh, India), and Chetan Arora (Indian Institute of Technology Delhi, India)

Bi-level Learning of Task-Specific Decoders for Joint Registration and One-Shot Medical Image Segmentation 11726
Xin Fan (Dalian University of Technology), Xiaolin Wang (Dalian University of Technology), Jiaxin Gao (Dalian University of Technology), Jia Wang (Dalian University of Technology), Zhongxuan Luo (Dalian University of Technology), and Risheng Liu (Dalian University of Technology)

PrPSeg: Universal Proposition Learning for Panoramic Renal Pathology Segmentation 11736
Ruining Deng (Vanderbilt University, USA), Quan Liu (Vanderbilt University, USA), Can Cui (Vanderbilt University, USA), Tianyuan Yao (Vanderbilt University, USA), Jialin Yue (Vanderbilt University, USA), Juming Xiong (Vanderbilt University, USA), Lining Yu (Vanderbilt University, USA), Yifei Wu (Vanderbilt University, USA), Mengmeng Yin (Vanderbilt University, USA), Yu Wang (Vanderbilt University Medical Center, USA), Shilin Zhao (Vanderbilt University Medical Center, USA), Yucheng Tang (NVIDIA Corp., USA), Haichun Yang (Vanderbilt University Medical Center, USA), and Yuankai Huo (Vanderbilt University, USA)

Versatile Medical Image Segmentation Learned from Multi-Source Datasets via Model Self-Disambiguation 11747
Xiaoyang Chen (University of Pennsylvania), Hao Zheng (University of Pennsylvania), Yuemeng Li (University of Pennsylvania), Yuncong Ma (University of Pennsylvania), Liang Ma (University of Pennsylvania), Hongming Li (University of Pennsylvania), and Yong Fan (University of Pennsylvania)

Masked Autoencoders for Microscopy are Scalable Learners of Cellular Biology 11757
Oren Kraus (Recursion, Canada), Kian Kenyon-Dean (Recursion, Canada), Saber Saberian (Recursion, Canada), Maryam Fallah (Recursion, Canada), Peter McLean (Recursion, USA), Jess Leung (Recursion, Canada), Vasudev Sharma (Recursion, Canada), Ayla Khan (Recursion, USA), Jia Balakrishnan (Recursion, USA), Safiye Celik (Recursion, USA), Dominique Beaini (Valence Labs, Canada), Maciej Sypetkowski (Valence Labs, Canada), Chi Vicky Cheng (Recursion, USA), Kristen Morse (Recursion, USA), Maureen Makes (Recursion, USA), Ben Mabey (Recursion, USA), and Berton Earnshaw (Recursion, USA)

EMCAD: Efficient Multi-scale Convolutional Attention Decoding for Medical Image Segmentation 11769
Md Mostafijur Rahman (The University of Texas at Austin, USA), Mustafa Munir (The University of Texas at Austin, USA), and Radu Marculescu (The University of Texas at Austin, USA)

Neural Underwater Scene Representation 11780
Yunkai Tang (Peking University, China), Chengxuan Zhu (Peking University, China), Renjie Wan (Hong Kong Baptist University, China), Chao Xu (Peking University, China), and Boxin Shi (Peking University, China)

Hearing Anything Anywhere 11790
Mason Long Wang (Stanford University), Ryosuke Sawata (Stanford University, SONY AI), Samuel Clarke (Stanford University), Ruohan Gao (Stanford University, University of Maryland, College Park), Shangzhe Wu (Stanford University), and Jiajun Wu (Stanford University)

VMINer: Versatile Multi-view Inverse Rendering with Near- and Far-field Light Sources 11800
Fan Fei (Peking University, China), Jiajun Tang (Peking University, China), Ping Tan (Hong Kong University of Science and Technology, China), and Boxin Shi (Peking University, China)

DiLiGenRT: A Photometric Stereo Dataset with Quantified Roughness and Translucency 11810
Heng Guo (Beijing University of Posts and Telecommunications), Jieji Ren (Shanghai Jiaotong University), Feishi Wang (Peking University), Boxin Shi (Peking University), Mingjun Ren (Shanghai Jiaotong University), and Yasuyuki Matsushita (Osaka University)

NeRSP: Neural 3D Reconstruction for Reflective Objects with Sparse Polarized Images 11821
Yufei Han (Beijing University of Posts and Telecommunications, China), Heng Guo (Beijing University of Posts and Telecommunications, China), Koki Fukai (Osaka University, Japan), Hiroaki Santo (Osaka University, Japan), Boxin Shi (Peking University, China), Fumio Okura (Osaka University, Japan), Zhanyu Ma (Beijing University of Posts and Telecommunications, China), and Yunpeng Jia (Beijing University of Posts and Telecommunications, China)

Differentiable Display Photometric Stereo 11831
Seokjun Choi (POSTECH, South Korea), Seungwoo Yoon (POSTECH, South Korea), Giljoo Nam (Meta Reality Labs, USA), Seungyong Lee (POSTECH, South Korea), and Seung-Hwan Baek (POSTECH, South Korea)

Bayesian Differentiable Physics for Cloth Digitalization 11841
Deshan Gong (University of Leeds, United Kingdom), Ningtao Mao (University of Leeds, United Kingdom), and He Wang (University College London, United Kingdom)

Atlantis: Enabling Underwater Depth Estimation with Stable Diffusion 11852
Fan Zhang (Beijing Institute of Technology), Shaodi You (University of Amsterdam), Yu Li (International Digital Economy Academy), and Ying Fu (Beijing Institute of Technology)

Sparse Views, Near Light: A Practical Paradigm for Uncalibrated Point-light Photometric Stereo 11862
Mohammed Brahimi (Technical University of Munich, Munich Center for Machine Learning), Bjoern Haefner (Technical University of Munich, Munich Center for Machine Learning, NVIDIA), Zhenzhang Ye (Technical University of Munich, Munich Center for Machine Learning), Bastian Goldluecke (University of Konstanz), and Daniel Cremers (Technical University of Munich, Munich Center for Machine Learning)

Diffusion Reflectance Map: Single-Image Stochastic Inverse Rendering of Illumination and Reflectance 11873
Yuto Enyo (Kyoto University, Japan) and Ko Nishino (Kyoto University, Japan)

Deep Single Image Camera Calibration by Heatmap Regression to Recover Fisheye Images Under Manhattan World Assumption 11884
Nobuhiko Wakai (Panasonic Holdings Corporation, Japan), Satoshi Sato (Panasonic Holdings Corporation, Japan), Yasunori Ishii (Panasonic Holdings Corporation, Japan), and Takayoshi Yamashita (Chubu University, Japan)

Physics-guided Shape-from-Template: Monocular Video Perception through Neural Surrogate Models 11895
David Stotko (University of Bonn, Germany), Nils Wandel (University of Bonn, Germany), and Reinhard Klein (University of Bonn, Germany)

Spin-UP: Spin Light for Natural Light Uncalibrated Photometric Stereo 11905
Zongrui Li (Nanyang Technological University), Zhan Lu (Nanyang Technological University), Haojie Yan (Zhejiang University), Boxin Shi (Peking University), Gang Pan (Zhejiang University), Qian Zheng (Zhejiang University), and Xudong Jiang (Nanyang Technological University)

Discontinuity-preserving Normal Integration with Auxiliary Edges 11915
Hyomin Kim (POSTECH), Yucheol Jung (POSTECH), and Seungyong Lee (POSTECH)

A Theory of Joint Light and Heat Transport for Lambertian Scenes 11924
Mani Ramanagopal (Carnegie Mellon University, USA), Sriram Narayanan (Carnegie Mellon University, USA), Aswin C. Sankaranarayanan (Carnegie Mellon University, USA), and Srinivasa G. Narasimhan (Carnegie Mellon University, USA)

IDGuard: Robust, General, Identity-centric POI Proactive Defense Against Face Editing Abuse 11934
Yunshu Dai (Sun Yat-sen University), Jianwei Fei (Nanjing University), and Fangjun Huang (Sun Yat-sen University)

Ungeneralizable Examples 11944
Jingwen Ye (National University of Singapore, Singapore) and Xinchao Wang (National University of Singapore, Singapore)

Distilled Datamodel with Reverse Gradient Matching 11954
Jingwen Ye (National University of Singapore, Singapore), Ruonan Yu (National University of Singapore, Singapore), Songhua Liu (National University of Singapore, Singapore), and Xinchao Wang (National University of Singapore, Singapore)

EditGuard: Versatile Image Watermarking for Tamper Localization and Copyright Protection ... 11964
Xuanyu Zhang (Peking University), Runyi Li (Peking University), Jiwen Yu (Peking University), Youmin Xu (Peking University), Weiqi Li (Peking University), and Jian Zhang (Peking University)

SocialCounterfactuals: Probing and Mitigating Intersectional Social Biases in Vision-Language Models with Counterfactual Examples 11975
Phillip Howard (Intel), Avinash Madasu (Intel), Tiep Le (Intel), Gustavo Lujan Moreno (Intel), Anahita Bhiwandiwalla (Intel), and Vasudev Lal (Intel)

FedAS: Bridging Inconsistency in Personalized Federated Learning 11986
Xiyuan Yang (Wuhan University, China), Wenke Huang (Wuhan University, China), and Mang Ye (Wuhan University, China)

FairRAG: Fair Human Generation via Fair Retrieval Augmentation 11996
Robik Shrestha (AWS AI Labs, USA), Yang Zou (Amazon AGI, USA), Qiuyu Chen (Amazon AGI, USA), Zhiheng Li (Amazon AGI, USA), Yusheng Xie (Amazon AGI, USA), and Siqi Deng (Amazon AGI, USA)

Self-Discovering Interpretable Diffusion Latent Directions for Responsible Text-to-Image Generation 12006
Hang Li (LMU Munich, Germany; Munich Center for Machine Learning, Germany; Siemens AG, Germany), Chengzhi Shen (Technical University of Munich, Germany), Philip Torr (University of Oxford, UK), Volker Tresp (LMU Munich, Germany; Munich Center for Machine Learning, Germany), and Jindong Gu (University of Oxford, UK)

ExMap: Leveraging Explainability Heatmaps for Unsupervised Group Robustness to Spurious Correlations 12017
Rwiddhi Chakraborty (UiT The Arctic University of Norway), Adrian Sletten (UiT The Arctic University of Norway), and Michael C. Kampffmeyer (UiT The Arctic University of Norway)

Data Valuation and Detections in Federated Learning 12027
Wenqian Li (National University of Singapore), Shuran Fu (NUS (Chongqing) Research Institution), Fengrui Zhang (Georgetown University), and Yan Pang (National University of Singapore)

Utility-Fairness Trade-Offs and How to Find Them 12037
Sepehr Dehdashtian (Michigan State University, USA), Bashir Sadeghi (Michigan State University, USA), and Vishnu Naresh Boddeti (Michigan State University, USA)

SimAC: A Simple Anti-Customization Method for Protecting Face Privacy against Text-to-Image Synthesis of Diffusion Models 12047
Feifei Wang (University of Science and Technology of China; AlibabaCloud), Zhentao Tan (Alibaba Cloud; University of Science and Technology of China), Tianyi Wei (University of Science and Technology of China), Yue Wu (Alibaba Cloud), and Qidong Huang (University of Science and Technology of China)

GLOW: Global Layout Aware Attacks on Object Detection 12057
Jun Bao (The State Key Laboratory of Blockchain and Data Security, China), Buyu Liu (Zhejiang University, China), Kui Ren (Zhejiang University, China), and Jun Yu (Hangzhou Dianzi University and Harbin Institute of Technology, China)

FADES: Fair Disentanglement with Sensitive Relevance 12067
Taeuk Jang (Purdue University, USA) and Xiaoqian Wang (Purdue University, USA)

Fair Federated Learning under Domain Skew with Local Consistency and Domain Diversity 12077
Yuhang Chen (Wuhan University, China), Wenke Huang (Wuhan University, China), and Mang Ye (Wuhan University, China)

WateRF: Robust Watermarks in Radiance Fields for Protection of Copyrights 12087
Youngdong Jang (Korea University, Republic of Korea), Dong In Lee (Korea University, Republic of Korea), MinHyuk Jang (Korea University, Republic of Korea), Jong Wook Kim (Korea University, Republic of Korea), Feng Yang (Google Reaserch, USA), and Sangpil Kim (Korea University, Republic of Korea)

FLHetBench: Benchmarking Device and State Heterogeneity in Federated Learning 12098
Junyuan Zhang (Beihang University, China), Shuang Zeng (The University of Hong Kong, China), Miao Zhang (New York University, USA), Runxi Wang (Beihang University, China), Feifei Wang (The University of Hong Kong, China), Yuyin Zhou (UC Santa Cruz, USA), Paul Pu Liang (Carnegie Mellon University, USA), and Liangqiong Qu (The University of Hong Kong, China)

An Upload-Efficient Scheme for Transferring Knowledge From a Server-Side Pre-trained Generator to Clients in Heterogeneous Federated Learning 12109
Jianqing Zhang (Shanghai Jiao Tong University, China), Yang Liu (Tsinghua University, China; Shanghai Artificial Intelligence Laboratory, China), Yang Hua (Queen's University Belfast, United Kingdom), and Jian Cao (Shanghai Jiao Tong University, China; Shanghai Key Laboratory of Trusted Data Circulation and Governance in Web3, China)

Privacy-Preserving Optics for Enhancing Protection in Face De-Identification 12120
Jhon Lopez (Universidad Industrial de Santander), Carlos Hinojosa (King Abdullah University of Science and Technology), Henry Arguello (Universidad Industrial de Santander), and Bernard Ghanem (King Abdullah University of Science and Technology)

A Stealthy Wrongdoer: Feature-Oriented Reconstruction Attack against Split Learning 12130
Xiaoyang Xu (Wuhan University, China), Mengda Yang (Wuhan University, China), Wenzhe Yi (Wuhan University, China), Ziang Li (Wuhan University, China), Juan Wang (Wuhan University, China), Hongxin Hu (University at Buffalo, USA), Yong Zhuang (Wuhan University, China), and Yaxin Liu (Wuhan University, China)

RCL: Reliable Continual Learning for Unified Failure Detection 12140
Fei Zhu (Centre for Artificial Intelligence and Robotics, HKISI-CAS), Zhen Cheng (State Key Laboratory of Multimodal Artificial Intelligence Systems, CASIA), Xu-Yao Zhang (State Key Laboratory of Multimodal Artificial Intelligence Systems, CASIA), Cheng-Lin Liu (State Key Laboratory of Multimodal Artificial Intelligence Systems, CASIA), and Zhaoxiang Zhang (State Key Laboratory of Multimodal Artificial Intelligence Systems, CASIA)

Global and Local Prompts Cooperation via Optimal Transport for Federated Learning 12151
Hongxia Li (ShanghaiTech University), Wei Huang (RIKEN Center for Advanced Intelligence Project), Jingya Wang (ShanghaiTech University), and Ye Shi (ShanghaiTech University)

Gaussian Shading: Provable Performance-Lossless Image Watermarking for Diffusion Models .. 12162
Zijin Yang (University of Science and Technology of China, China), Kai Zeng (University of Science and Technology of China, China), Kejiang Chen (University of Science and Technology of China, China), Han Fang (National University of Singapore, Singapore), Weiming Zhang (University of Science and Technology of China, China), and Nenghai Yu (University of Science and Technology of China, China)

Explaining CLIP's Performance Disparities on Data from Blind/Low Vision Users 12172
Daniela Massiceti (Microsoft Research, Australia), Camilla Longden (Microsoft Research, UK), Agnieszka Slowik (Microsoft Research, UK), Samuel Wills (The World Bank, Australia), Martin Grayson (Microsoft Research, UK), and Cecily Morrison (Microsoft Research, UK)

Model Inversion Robustness: Can Transfer Learning Help? 12183
Sy-Tuyen Ho (Singapore University of Technology and Design, Singapore), Koh Jun Hao (Singapore University of Technology and Design, Singapore), Keshigeyan Chandrasegaran (Stanford University, USA), Ngoc-Bao Nguyen (Singapore University of Technology and Design, Singapore), and Ngai-Man Cheung (Singapore University of Technology and Design, Singapore)

Make Me a BNN: A Simple Strategy for Estimating Bayesian Uncertainty from Pre-trained Models 12194
Gianni Franchi (U2IS, ENSTA Paris, Institut Polytechnique de Paris, France), Olivier Laurent (U2IS, ENSTA Paris, Institut Polytechnique de Paris, France; SATIE, Paris Saclay University), Maxence Leguery (U2IS, ENSTA Paris, Institut Polytechnique de Paris, France), Andrei Bursuc (Valeo.ai), Andrea Pilzer (NVIDIA), and Angela Yao (National University of Singapore)

Validating Privacy-Preserving Face Recognition under a Minimum Assumption 12205
Hui Zhang (Anhui University, China), Xingbo Dong (Anhui University, China), YenLung Lai (Anhui University, China), Ying Zhou (Anhui University, China), Xiaoyan Zhang (Anhui University, China), Xingguo Lv (Anhui University, China), Zhe Jin (Anhui University, China), and Xuejun Li (Anhui University, China)

Re-thinking Data Availability Attacks Against Deep Neural Networks 12215
Bin Fang (Shanghai Jiao Tong University), Bo Li (Youtu Lab Tencent), Shuang Wu (Youtu Lab Tencent), Shouhong Ding (Youtu Lab Tencent), Ran Yi (Shanghai Jiao Tong University), and Lizhuang Ma (Shanghai Jiao Tong University)

OpenBias: Open-set Bias Detection in Text-to-Image Generative Models 12225
Moreno D'Incà (University of Trento, Italy), Elia Peruzzo (University of Trento, Italy), Massimiliano Mancini (University of Trento, Italy), Dejia Xu (University of Texas at Austin, USA), Vidit Goel (Picsart Inc, USA), Xingqian Xu (University of Texas at Austin, USA; Picsart Inc, USA), Zhangyang Wang (University of Texas at Austin, USA; Picsart Inc, USA), Humphrey Shi (Georgia Tech, USA; Picsart Inc, USA; UIUC, USA), and Nicu Sebe (University of Trento, Italy)

In-distribution Public Data Synthesis with Diffusion Models for Differentially Private Image Classification 12236
Jinseong Park (Seoul National University, Republic of Korea), Yujin Choi (Seoul National University, Republic of Korea), and Jaewook Lee (Seoul National University, Republic of Korea)

Leak and Learn: An Attacker's Cookbook to Train Using Leaked Data from Federated Learning 12247
Joshua C. Zhao (Purdue University), Ahaan Dabholkar (Purdue University), Atul Sharma (Purdue University), and Saurabh Bagchi (Purdue University)

Countering Personalized Text-to-Image Generation with Influence Watermarks 12257
Hanwen Liu (Peking University, China), Zhicheng Sun (Peking University, China), and Yadong Mu (Peking University, China)

Fair-VPT: Fair Visual Prompt Tuning for Image Classification 12268
Sungho Park (Yonsei University, Republic of Korea) and Hyeran Byun (Yonsei University, Republic of Korea)

Relaxed Contrastive Learning for Federated Learning 12279
Seonguk Seo (Seoul National University), Jinkyu Kim (Seoul National University), Geeho Kim (Seoul National University), and Bohyung Han (Seoul National University)

FairCLIP: Harnessing Fairness in Vision-Language Learning 12289
Yan Luo (Harvard University, USA), Min Shi (Harvard University, USA), Muhammad Osama Khan (New York University, USA), Muhammad Muneeb Afzal (New York University, USA), Hao Huang (New York University, USA), Shuaihang Yuan (New York University), Yu Tian (Harvard University, USA), Luo Song (Harvard University, USA), Ava Kouhana (Harvard University, USA), Tobias Elze (Harvard University, USA), Yi Fang (New York University, USA), and Mengyu Wang (Harvard University, USA)

Steganographic Passport: An Owner and User Verifiable Credential for Deep Model IP Protection Without Retraining 12302
Qi Cui (Nanyang Technological University), Ruohan Meng (Nanyang Technological University), Chaohui Xu (Nanyang Technological University), and Chip-Hong Chang (Nanyang Technological University)

Adaptive Hyper-graph Aggregation for Modality-Agnostic Federated Learning 12312
Fan Qi (Tianjin University of Technology, Tianjin, China) and Shuai Li (Tianjin University of Technology, Tianjin, China)

Navigate Beyond Shortcuts: Debiased Learning Through the Lens of Neural Collapse 12322
Yining Wang (Fudan University, China), Junjie Sun (Fudan University, China), Chenyue Wang (Fudan University, China), Mi Zhang (Fudan University, China), and Min Yang (Fudan University, China)

Enhancing Intrinsic Features for Debiasing via Investigating Class-Discerning Common Attributes in Bias-Contrastive Pair 12332
Jeonghoon Park (Korea Advanced Institute of Science and Technology (KAIST), South Korea), Chaeyeon Chung (Korea Advanced Institute of Science and Technology (KAIST), South Korea), and Jaegul Choo (Korea Advanced Institute of Science and Technology (KAIST), South Korea)

Device-Wise Federated Network Pruning 12342
Shangqian Gao (University of Pittsburgh, USA), Junyi Li (University of Maryland College Park, USA), Zeyu Zhang (University of Arizona, USA), Yanfu Zhang (College of William and Mary, USA), Weidong Cai (University of Sydney, Australia), and Heng Huang (University of Maryland College Park, USA)

All Rivers Run to the Sea: Private Learning with Asymmetric Flows 12353
Yue Niu (University of Southern California), Ramy E. Ali (Samsung), Saurav Prakash (University of Illinois Urbana-Champaign), and Salman Avestimehr (University of Southern California)

VA3: Virtually Assured Amplification Attack on Probabilistic Copyright Protection for Text-to-Image Generative Models 12363
Xiang Li (National University of Singapore, Singapore), Qianli Shen (National University of Singapore, Singapore), and Kenji Kawaguchi (National University of Singapore, Singapore)

CPR: Retrieval Augmented Generation for Copyright Protection 12374
Aditya Golatkar (Amazon Web Services), Alessandro Achille (Amazon Web Services), Luca Zancato (Amazon Web Services), Yu-Xiang Wang (Amazon Web Services), Ashwin Swaminathan (Amazon Web Services), and Stefano Soatto (Amazon Web Services)

Communication-Efficient Federated Learning with Accelerated Client Gradient 12385
Geeho Kim (ECE, Seoul National University), Jinkyu Kim (ECE, Seoul National University), and Bohyung Han (ECE & IPAI, Seoul National University)

Self-supervised Debiasing Using Low Rank Regularization 12395
Geon Yeong Park (KAIST, South Korea), Chanyong Jung (KAIST, South Korea), Sangmin Lee (KAIST, South Korea), Jong Chul Ye (KAIST, South Korea), and Sang Wan Lee (KAIST, South Korea)

Facial Identity Anonymization via Intrinsic and Extrinsic Attention Distraction 12406
Zhenzhong Kuang (Hangzhou Dianzi University, China), Xiaochen Yang (Hangzhou Dianzi University, China), Yingjie Shen (Hangzhou Dianzi University, China), Chao Hu (Hangzhou Dianzi University, China), and Jun Yu (Hangzhou Dianzi University, Harbin Institute of Technology (Shenzhen), China)

Collaborative Learning of Anomalies with Privacy (CLAP) for Unsupervised Video Anomaly Detection: A New Baseline 12416
Anas Al-lahham (Mohamed bin Zayed University of Artificial Intelligence), Muhammad Zaigham Zaheer (Mohamed bin Zayed University of Artificial Intelligence), Nurbek Tastan (Mohamed bin Zayed University of Artificial Intelligence), and Karthik Nandakumar (Mohamed bin Zayed University of Artificial Intelligence)

Label-Efficient Group Robustness via Out-of-Distribution Concept Curation 12426
Yiwei Yang (University of Washington, USA), Anthony Z. Liu (University of Michigan, USA), Robert Wolfe (University of Washington), Aylin Caliskan (University of Washington, USA), and Bill Howe (University of Washington, USA)

Long-Tailed Anomaly Detection with Learnable Class Names 12435
Chih-Hui Ho (University of California San Diego), Kuan-Chuan Peng (Mitsubishi Electric Research Laboratories (MERL)), and Nuno Vasconcelos (University of California San Diego)

Robust Emotion Recognition in Context Debiasing 12447
Dingkang Yang (Fudan University), Kun Yang (Fudan University), Mingcheng Li (Fudan University), Shunli Wang (Fudan University), Shuaibing Wang (Fudan University), and Lihua Zhang (Fudan University)

Correlation-Decoupled Knowledge Distillation for Multimodal Sentiment Analysis with Incomplete Modalities 12458
Mingcheng Li (Fudan University), Dingkang Yang (Fudan University), Xiao Zhao (Fudan University), Shuaibing Wang (Fudan University), Yan Wang (Fudan University), Kun Yang (Fudan University), Mingyang Sun (Fudan University), Dongliang Kou (Fudan University), Ziyun Qian (Fudan University), and Lihua Zhang (Fudan University)

An Edit Friendly DDPM Noise Space: Inversion and Manipulations 12469
Inbar Huberman-Spiegelglas (Technion – Israel Institute of Technology, Israel), Vladimir Kulikov (Technion – Israel Institute of Technology, Israel), and Tomer Michaeli (Technion – Israel Institute of Technology, Israel)

SleepVST: Sleep Staging from Near-Infrared Video Signals using Pre-Trained Transformers 12479
Jonathan F. Carter (University of Oxford), João Jorge (Oxehealth Ltd., Oxford), Oliver Gibson (Oxehealth Ltd., Oxford), and Lionel Tarassenko (University of Oxford)

AM-RADIO: Agglomerative Vision Foundation Model Reduce All Domains Into One 12490
Mike Ranzinger (NVIDIA), Greg Heinrich (NVIDIA), Jan Kautz (NVIDIA), and Pavlo Molchanov (NVIDIA)

Towards Language-Driven Video Inpainting via Multimodal Large Language Models 12501
Jianzong Wu (Peking University, China), Xiangtai Li (Nanyang Technological University, Singapore), Chenyang Si (Nanyang Technological University, Singapore), Shangchen Zhou (Nanyang Technological University, Singapore), Jingkang Yang (Nanyang Technological University, Singapore), Jiangning Zhang (Zhejiang University, China), Yining Li (Shanghai AI Laboratory, China), Kai Chen (Shanghai AI Laboratory, China), Yunhai Tong (Peking University), Ziwei Liu (Nanyang Technological University, Singapore), and Chen Change Loy (Nanyang Technological University, Singapore)

FedSOL: Stabilized Orthogonal Learning with Proximal Restrictions in Federated Learning 12512
Gihun Lee (KAIST), Minchan Jeong (KAIST), Sangmook Kim (UBC), Jaehoon Oh (Samsung Advanced Institute of Technology), and Se-Young Yun (KAIST)

UnionFormer: Unified-Learning Transformer with Multi-View Representation for Image Manipulation Detection and Localization 12523
Shuaibo Li (Beijing University of Technology, China & Chinese Academy of Sciences, China), Wei Ma (Beijing University of Technology, China), Jianwei Guo (Chinese Academy of Sciences, China), Shibiao Xu (Beijing University of Posts and Telecommunications, China), Benchong Li (Beijing University of Technology, China), and Xiaopeng Zhang (Chinese Academy of Sciences, China)

Motion Blur Decomposition with Cross-shutter Guidance 12534
Xiang Ji (The University of Tokyo, Japan), Haiyang Jiang (The University of Tokyo, Japan), and Yinqiang Zheng (The University of Tokyo, Japan)

SNIDA: Unlocking Few-Shot Object Detection with Non-linear Semantic Decoupling Augmentation 12544
Yanjie Wang (National Key Laboratory of Multispectral Information Intelligent Processing Technology, Huazhong University of Science and Technology, China), Xu Zou (National Key Laboratory of Multispectral Information Intelligent Processing Technology, Huazhong University of Science and Technology, China), Luxin Yan (National Key Laboratory of Multispectral Information Intelligent Processing Technology, Huazhong University of Science and Technology, China), Sheng Zhong (National Key Laboratory of Multispectral Information Intelligent Processing Technology, Huazhong University of Science and Technology,China), and Jiahuan Zhou (Wangxuan Institute of Computer Technology, Peking University, China)

Rapid 3D Model Generation with Intuitive 3D Input 12554
Tianrun Chen (Zhejiang University, China; KOKONI3D, Moxin (Huzhou) Technology Co., China), Chaotao Ding (Huzhou University, China), Shangzhan Zhang (Zhejiang University, China), Chunan Yu (Huzhou University, China), Ying Zang (Huzhou University, China), Zejian Li (Zhejiang University, China), Sida Peng (Zhejiang University, China), and Lingyun Sun (Zhejiang University, China)

SketchINR: A First Look into Sketches as Implicit Neural Representations 12565
Hmrishav Bandyopadhyay (University of Surrey), Ayan Kumar Bhunia (University of Surrey), Pinaki Nath Chowdhury (University of Surrey), Aneeshan Sain (University of Surrey), Tao Xiang (University of Surrey, iFlyTek-Surrey Joint Research Centre on Artificial Intelligence), Timothy Hospedales (University of Edinburgh), and Yi-Zhe Song (University of Surrey, iFlyTek-Surrey Joint Research Centre on Artificial Intelligence)

ERMVP: Communication-Efficient and Collaboration-Robust Multi-Vehicle Perception in Challenging Environments 12575
Jingyu Zhang (Fudan University, China), Kun Yang (Fudan University, China), Yilei Wang (Fudan University, China), Hanqi Wang (Fudan University, China), Peng Sun (Duke Kunshan University, China), and Liang Song (Fudan University, China)

DiaLoc: An Iterative Approach to Embodied Dialog Localization 12585
Chao Zhang (Toshiba Europe Ltd), Mohan Li (Toshiba Europe Ltd), Ignas Budvytis (Toshiba Europe Ltd), and Stephan Liwicki (Toshiba Europe Ltd)

WildlifeMapper: Aerial Image Analysis for Multi-Species Detection and Identification 12594
Satish Kumar (University of California Santa Barbara), Bowen Zhang (University of California Santa Barbara), Chandrakanth Gudavalli (University of California Santa Barbara), Connor Levenson (University of California Santa Barbara), Lacey Hughey (Smithsonian National Zoo and Conservation Biology Institute), Jared A. Stabach (Smithsonian National Zoo and Conservation Biology Institute), Irene Amoke (Kenya Wildlife Trust), Gordon Ojwang (University of Groningen), Joseph Mukeka (Wildlife Research and Training Institute), Stephen Mwiu (Wildlife Research and Training Institute), Joseph Ogutu (University of Hohenheim), Howard Frederick (Tanzania Wildlife Research Institute), and B.S. Manjunath (University of California Santa Barbara)

Harnessing Meta-Learning for Improving Full-Frame Video Stabilization 12605
Muhammad Kashif Ali (Hanyang University), Eun Woo Im (Hanyang University), Dongjin Kim (Hanyang University), and Tae Hyun Kim (Hanyang University)

De-confounded Data-free Knowledge Distillation for Handling Distribution Shifts 12615
Yuzheng Wang (Fudan University, China), Dingkang Yang (Fudan University, China), Zhaoyu Chen (Fudan University, China), Yang Liu (Fudan University, China), Siao Liu (Fudan University, China), Wenqiang Zhang (Fudan University, China), Lihua Zhang (Fudan University, China), and Lizhe Qi (Fudan University, China)

Day-Night Cross-domain Vehicle Re-identification 12626
Hongchao Li (Anhui Normal University, China), Jingong Chen (Anhui Normal University, China), Aihua Zheng (Anhui University, China), Yong Wu (Anhui Normal University, China), and Yonglong Luo (Anhui Normal University, China)

Brush2Prompt: Contextual Prompt Generator for Object Inpainting 12636
Mang Tik Chiu (UIUC, Adobe), Yuqian Zhou (Adobe), Lingzhi Zhang (Adobe), Zhe Lin (Adobe), Connelly Barnes (Adobe), Sohrab Amirghodsi (Adobe), Eli Shechtman (Adobe), and Humphrey Shi (University of Oregon)

Cloud-Device Collaborative Learning for Multimodal Large Language Models 12646
Guanqun Wang (Peking University), Jiaming Liu (Peking University), Chenxuan Li (Peking University), Yuan Zhang (Peking University), Junpeng Ma (Peking University), Xinyu Wei (Peking University), Kevin Zhang (Peking University), Maurice Chong (Peking University, KNQ.AI), Renrui Zhang (The Chinese University of Hong Kong), Yijiang Liu (Nanjing University), and Shanghang Zhang (Peking University)

Making Visual Sense of Oracle Bones for You and Me 12656
Runqi Qiao (Beijing University of Posts and Telecommunications, China), Lan Yang (Beijing University of Posts and Telecommunications, China), Kaiyue Pang (University of Surrey, UK), and Honggang Zhang (Beijing University of Posts and Telecommunications, China)

Boosting Object Detection with Zero-Shot Day-Night Domain Adaptation 12666
Zhipeng Du (King's College London), Miaojing Shi (Tongji University), and Jiankang Deng (Imperial College London)

InNeRF360: Text-Guided 3D-Consistent Object Inpainting on 360-degree Neural Radiance Fields 12677
Dongqing Wang (EPFL, Switzerland), Tong Zhang (EPFL, Switzerland), Alaa Abboud (EPFL, Switzerland), and Sabine Süsstrunk (EPFL, Switzerland)

Language Models as Black-Box Optimizers for Vision-Language Models 12687
Shihong Liu (Carnegie Mellon University), Samuel Yu (Carnegie Mellon University), Zhiqiu Lin (Carnegie Mellon University), Deepak Pathak (Carnegie Mellon University), and Deva Ramanan (Carnegie Mellon University)

Mind Marginal Non-Crack Regions: Clustering-Inspired Representation Learning for Crack Segmentation 12698
Zhuangzhuang Chen (Shenzhen Univeristy, China), Zhuonan Lai (Shenzhen Univeristy, China), Jie Chen (Shenzhen Univeristy, China), and Jianqiang Li (Shenzhen Univeristy, China)

InstructDiffusion: A Generalist Modeling Interface for Vision Tasks 12709
Zigang Geng (University of Science and Technology of China, China), Binxin Yang (University of Science and Technology of China, China), Tiankai Hang (Southeast University, China), Chen Li (Xi'an Jiaotong University, China), Shuyang Gu (Microsoft Research, China), Ting Zhang (Beijing Normal University, China), Jianmin Bao (Microsoft Research, China), Zheng Zhang (Microsoft Research, China), Houqiang Li (University of Science and Technology of China, China), Han Hu (Microsoft Research, China), Dong Chen (Microsoft Research, China), and Baining Guo (Microsoft Research, China)

Desigen: A Pipeline for Controllable Design Template Generation 12721
Haohan Weng (South China University of Technology, China), Danqing Huang (Microsoft, China), Yu Qiao (Central South University, China), Zheng Hu (Central South University, China), Chin-Yew Lin (Microsoft, China), Tong Zhang (South China University of Technology, China), and C. L. Philip Chen (South China University of Technology, China)

Physical Backdoor: Towards Temperature-based Backdoor Attacks in the Physical World 12733
Wen Yin (Huazhong University of Science and Technology, China), Jian Lou (Zhejiang University, China), Pan Zhou (Huazhong University of Science and Technology, China), Yulai Xie (Huazhong University of Science and Technology, China), Dan Feng (Huazhong University of Science and Technology, China), Yuhua Sun (Huazhong University of Science and Technology, China), Tailai Zhang (Huazhong University of Science and Technology, China), and Lichao Sun (Lehigh University, USA)

Behind the Veil: Enhanced Indoor 3D Scene Reconstruction with Occluded Surfaces Completion 12744
Su Sun (Purdue University, USA), Cheng Zhao (Bosch Research North America, USA), Yuliang Guo (Bosch Research North America, USA), Ruoyu Wang (Bosch Research North America, USA), Xinyu Huang (Bosch Research North America, USA), Yingjie Victor Chen (Purdue University, USA), and Liu Ren (Bosch Research North America, USA)

EarthLoc: Astronaut Photography Localization by Indexing Earth from Space 12754
Gabriele Berton (Polytechnic of Turin, Italy), Alex Stoken (NASA Johnson Space Center, USA), Barbara Caputo (Polytechnic of Turin, Italy), and Carlo Masone (Polytechnic of Turin, Italy)

DiffForensics: Leveraging Diffusion Prior to Image Forgery Detection and Localization 12765
Zeqin Yu (Sun Yat-sen University, China), Jiangqun Ni (Sun Yat-sen University, China; Peng Cheng Laboratory, China), Yuzhen Lin (Shenzhen University, China), Haoyi Deng (Shenzhen University, China), and Bin Li (Shenzhen University, China)

MuseChat: A Conversational Music Recommendation System for Videos 12775
Zhikang Dong (Stony Brook University, USA), Xiulong Liu (University of Washington, USA), Bin Chen (Bytedance.com, USA), Pawel Polak (Stony Brook University, USA), and Peng Zhang (Bytedance.com, USA)

The Unreasonable Effectiveness of Pre-Trained Features for Camera Pose Refinement 12786
Gabriele Trivigno (Politecnico di Torino, Italy), Carlo Masone (Politecnico di Torino, Italy), Barbara Caputo (Politecnico di Torino, Italy), and Torsten Sattler (Czech Technical University in Prague, Czechia)

Blind Image Quality Assessment Based on Geometric Order Learning 12799
Nyeong-Ho Shin (Korea University), Seon-Ho Lee (Korea University), and Chang-Su Kim (Korea University)

CrowdDiff: Multi-hypothesis Crowd Density Estimation using Diffusion Models 12809
Yasiru Ranasinghe (Johns Hopkins University, USA), Nithin Gopalakrishnan Nair (Johns Hopkins University, USA), Wele Gedara Chaminda Bandara (Johns Hopkins University, USA), and Vishal M. Patel (Johns Hopkins University, USA)

Towards Efficient Replay in Federated Incremental Learning 12820
Yichen Li (Huazhong University of Science and Technology), Qunwei Li (Ant Group), Haozhao Wang (Huazhong University of Science and Technology), Ruixuan Li (Huazhong University of Science and Technology), Wenliang Zhong (Ant Group), and Guannan Zhang (Ant Group)

MART: Masked Affective RepresenTation Learning via Masked Temporal Distribution Distillation 12830
Zhicheng Zhang (VCIP & TMCC & DISSec, Nankai University,Nankai International Advanced Research Institute (SHENZHEN· FUTIAN)), Pancheng Zhao (VCIP & TMCC & DISSec, Nankai University,Nankai International Advanced Research Institute (SHENZHEN· FUTIAN)), Eunil Park (Sungkyunkwan University), and Jufeng Yang (VCIP & TMCC & DISSec, Nankai University,Nankai International Advanced Research Institute (SHENZHEN· FUTIAN))

PolarRec: Improving Radio Interferometric Data Reconstruction Using Polar Coordinates 12841
Ruoqi Wang (The Hong Kong University of Science and Technology (Guangzhou)), Zhuoyang Chen (The Hong Kong University of Science and Technology (Guangzhou)), Jiayi Zhu (The Hong Kong University of Science and Technology (Guangzhou)), Qiong Luo (The Hong Kong University of Science and Technology (Guangzhou) & The Hong Kong University of Science and Technology), and Feng Wang (Guangzhou University)

Constrained Layout Generation with Factor Graphs 12851
Mohammed Haroon Dupty (National University of Singapore), Yanfei Dong (National University of Singapore, PayPal), Sicong Leng (Singapore University of Technology and Design), Guoji Fu (National University of Singapore), Yong Liang Goh (National University of Singapore), Wei Lu (Singapore University of Technology and Design), and Wee Sun Lee (National University of Singapore)

Visual In-Context Prompting 12861
Feng Li (The Hong Kong University of Science and Technology), Qing Jiang (South China University of Science and Technology), Hao Zhang (The Hong Kong University of Science and Technology), Tianhe Ren (International Digital Economy Academy), Shilong Liu (Tsinghua University), Xueyan Zou (University of Wisconsin-Madison), Huaizhe Xu (The Hong Kong University of Science and Technology), Hongyang Li (South China University of Science and Technology), Jianwei Yang (Microsoft Research), Chunyuan Li (Microsoft Research), Lei Zhang (International Digital Economy Academy), and Jianfeng Gao (Microsoft Research)

Traceable Federated Continual Learning 12872
Qiang Wang (Beijing University of Posts and Telecommunications, China), Bingyan Liu (Beijing University of Posts and Telecommunications, China), and Yawen Li (Beijing University of Posts and Telecommunications, China)

Interactive Continual Learning: Fast and Slow Thinking 12882
Biqing Qi (Harbin Institute of Technology; Tsinghua University; Frontis.AI), Xinquan Chen (Harbin Institute of Technology), Junqi Gao (Harbin Institute of Technology), Dong Li (Harbin Institute of Technology), Jianxing Liu (Harbin Institute of Technology), Ligang Wu (Harbin Institute of Technology), and Bowen Zhou (Harbin Institute of Technology; Tsinghua University; Frontis.AI)

PIGEON: Predicting Image Geolocations 12893
Lukas Haas (Stanford University, USA), Michal Skreta (Stanford University, USA), Silas Alberti (Stanford University, USA), and Chelsea Finn (Stanford University, USA)

LQMFormer: Language-aware Query Mask Transformer for Referring Image Segmentation 12903
Nisarg A. Shah (Johns Hopkins University, USA), Vibashan VS (Johns Hopkins University, USA), and Vishal M. Patel (Johns Hopkins University, USA)

ViP-LLaVA: Making Large Multimodal Models Understand Arbitrary Visual Prompts 12914
Mu Cai (University of Wisconsin-Madison, USA), Haotian Liu (University of Wisconsin-Madison, USA), Siva Karthik Mustikovela (Cruise LLC, USA), Gregory P. Meyer (Cruise LLC, USA), Yuning Chai (Cruise LLC, USA), Dennis Park (Cruise LLC, USA), and Yong Jae Lee (University of Wisconsin-Madison, USA)

DePT: Decoupled Prompt Tuning 12924
Ji Zhang (University of Electronic Science and Technology of China (UESTC), China), Shihan Wu (University of Electronic Science and Technology of China (UESTC), China), Lianli Gao (Shenzhen Institute for Advanced Study, UESTC, China), Heng Tao Shen (Tongji University, China), and Jingkuan Song (Tongji University, China)

Grounded Question-Answering in Long Egocentric Videos 12934
Shangzhe Di (CMIC, Shanghai Jiao Tong University, China) and Weidi Xie (CMIC, Shanghai Jiao Tong University, China; Shanghai AI Lab, China)

HalluciDoctor: Mitigating Hallucinatory Toxicity in Visual Instruction Data 12944
Qifan Yu (Zhejiang University, China), Juncheng Li (Zhejiang University, China), Longhui Wei (Huawei Cloud, China), Liang Pang (Chinese Academy of Sciences, China), Wentao Ye (Zhejiang University, China), Bosheng Qin (Zhejiang University, China), Siliang Tang (Zhejiang University, China), Qi Tian (Huawei Cloud, China), and Yueting Zhuang (Zhejiang University)

ViTamin: Designing Scalable Vision Models in the Vision-Language Era 12954
Jieneng Chen (Johns Hopkins University), Qihang Yu (ByteDance), Xiaohui Shen (ByteDance), Alan Yuille (Johns Hopkins University), and Liang-Chieh Chen (ByteDance)

The Manga Whisperer: Automatically Generating Transcriptions for Comics 12967
Ragav Sachdeva (University of Oxford, United Kingdom) and Andrew Zisserman (University of Oxford, United Kingdom)

Learning to Localize Objects Improves Spatial Reasoning in Visual-LLMs 12977
Kanchana Ranasinghe (Stony Brook University), Satya Narayan Shukla (Meta), Omid Poursaeed (Meta), Michael S. Ryoo (Stony Brook University), and Tsung-Yu Lin (Meta)

The Neglected Tails in Vision-Language Models 12988
Shubham Parashar (Texas A&M University, USA), Zhiqiu Lin (Carnegie Mellon University, USA), Tian Liu (Texas A&M University, USA), Xiangjue Dong (Texas A&M University, USA), Yanan Li (Zhejiang Lab), Deva Ramanan (Carnegie Mellon University, USA), James Caverlee (Texas A&M University, USA), and Shu Kong (Texas A&M University and University of Macau)

Unveiling Parts Beyond Objects: Towards Finer-Granularity Referring Expression Segmentation 12998
Wenxuan Wang (Institute of Automation, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China; Beijing Academy of Artificial Intelligence, China), Tongtian Yue (Institute of Automation, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), Yisi Zhang (University of Science and Technology Beijing, China), Longteng Guo (Institute of Automation, Chinese Academy of Sciences, China), Xingjian He (Institute of Automation, Chinese Academy of Sciences, China), Xinlong Wang (Beijing Academy of Artificial Intelligence, China), and Jing Liu (Institute of Automation, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China)

GLaMM: Pixel Grounding Large Multimodal Model 13009
Hanoona Rasheed (Mohamed bin Zayed University of Artificial Intelligence, UAE), Muhammad Maaz (Mohamed bin Zayed University of Artificial Intelligence, UAE), Sahal Shaji (Mohamed bin Zayed University of Artificial Intelligence, UAE), Abdelrahman Shaker (Mohamed bin Zayed University of Artificial Intelligence, UAE), Salman Khan (Mohamed bin Zayed University of Artificial Intelligence, UAE; Australian National University, Australia), Hisham Cholakkal (Mohamed bin Zayed University of Artificial Intelligence, UAE), Rao M. Anwer (Mohamed bin Zayed University of Artificial Intelligence, UAE; Aalto University, Finland), Eric Xing (Mohamed bin Zayed University of Artificial Intelligence, UAE; Carnegie Mellon University, USA), Ming-Hsuan Yang (University of California - Merced, USA; Google Research, USA), and Fahad S. Khan (Mohamed bin Zayed University of Artificial Intelligence, UAE; Linköping University, Sweden)

Alpha-CLIP: A CLIP Model Focusing on Wherever You Want 13019
Zeyi Sun (Shanghai Jiao Tong University), Ye Fang (Fudan University), Tong Wu (The Chinese University of Hong Kong), Pan Zhang (Shanghai Artificial Intelligence Laboratory), Yuhang Zang (Shanghai Artificial Intelligence Laboratory), Shu Kong (University of Macau), Yuanjun Xiong (MThreads, Inc.), Dahua Lin (The Chinese University of Hong Kong), and Jiaqi Wang (Shanghai Artificial Intelligence Laboratory)

Pixel-Aligned Language Model 13030
Jiarui Xu (UCSD, USA), Xingyi Zhou (Google Research, USA), Shen Yan (Google Research, USA), Xiuye Gu (Google Research, USA), Anurag Arnab (Google Research, France), Chen Sun (Google Research, USA), Xiaolong Wang (UCSD, USA), and Cordelia Schmid (Google Research, USA)

mPLUG-Owl2: Revolutionizing Multi-modal Large Language Model with Modality Collaboration 13040
Qinghao Ye (Alibaba Group, China), Haiyang Xu (Alibaba Group, China), Jiabo Ye (Alibaba Group, China), Ming Yan (Alibaba Group, China), Anwen Hu (Alibaba Group, China), Haowei Liu (Alibaba Group, China), Qi Qian (Alibaba Group, China), Ji Zhang (Alibaba Group, China), and Fei Huang (Alibaba Group, China)

SNIFFER: Multimodal Large Language Model for Explainable Out-of-Context Misinformation Detection 13052
Peng Qi (National University of Singapore, Singapore), Zehong Yan (National University of Singapore, Singapore), Wynne Hsu (National University of Singapore, Singapore), and Mong Li Lee (National University of Singapore, Singapore)

Towards CLIP-driven Language-free 3D Visual Grounding via 2D-3D Relational Enhancement and Consistency 13063
Yuqi Zhang (Sichuan University, China), Han Luo (Sichuan University, China), and Yinjie Lei (Sichuan University, China)

SC-Tune: Unleashing Self-Consistent Referential Comprehension in Large Vision Language Models 13073
Tongtian Yue (CASIA, China), Jie Cheng (CASIA, China), Longteng Guo (CASIA, China), Xingyuan Dai (CASIA, China), Zijia Zhao (CASIA, China), Xingjian He (CASIA, China), Gang Xiong (CASIA, China), Yisheng Lv (CASIA, China), and Jing Liu (CASIA, China)

V: Guided Visual Search as a Core Mechanism in Multimodal LLMs 13084
Penghao Wu (UC San Diego) and Saining Xie (New York University)

Improved Visual Grounding through Self-Consistent Explanations 13095
Ruozhen He (Rice University), Paola Cascante-Bonilla (Rice University), Ziyan Yang (Rice University), Alexander C. Berg (University of California, Irvine), and Vicente Ordonez (Rice University)

Distilling Vision-Language Models on Millions of Videos 13106
Yue Zhao (University of Texas, Austin), Long Zhao (Google Research), Xingyi Zhou (Google Research), Jialin Wu (Google Research), Chun-Te Chu (Google Research), Hui Miao (Google Research), Florian Schroff (Google Research), Hartwig Adam (Google Research), Ting Liu (Google Research), Boqing Gong (Google Research), Philipp Krahenbuhl (University of Texas, Austin), and Liangzhe Yuan (Google Research)

Separating the "Chirp" from the "Chat": Self-supervised Visual Grounding of Sound and Language 13117
Mark Hamilton (MIT), Andrew Zisserman (Oxford), John R. Hershey (Google), and William T. Freeman (MIT)

Referring Image Editing: Object-level Image Editing via Referring Expressions 13128
Chang Liu (Fudan University, China), Xiangtai Li (Nanyang Technological University, Singapore), and Henghui Ding (Fudan University, China)

Vision-and-Language Navigation via Causal Learning 13139
Liuyi Wang (Tongji University, China), Zongtao He (Tongji University, China), Ronghao Dang (Tongji University, China), Mengjiao Shen (Tongji University, China), Chengju Liu (Tongji University, China), and Qijun Chen (Tongji University, China)

VISTA-LLAMA: Reducing Hallucination in Video Language Models via Equal Distance to Visual Tokens 13151
Fan Ma (Zhejiang University), Xiaojie Jin (Bytedance Inc.), Heng Wang (Bytedance Inc.), Yuchen Xian (Zhejiang University), Jiashi Feng (Bytedance Inc.), and Yi Yang (Zhejiang University)

Ranking Distillation for Open-Ended Video Question Answering with Insufficient Labels 13161
Tianming Liang (Sun Yat-sen University), Chaolei Tan (Sun Yat-sen University), Beihao Xia (Huazhong University of Science and Technology), Wei-Shi Zheng (Sun Yat-sen University), and Jian-Fang Hu (Sun Yat-sen University)

CLIP as RNN: Segment Countless Visual Concepts without Training Endeavor 13171
Shuyang Sun (University of Oxford), Runjia Li (University of Oxford), Philip Torr (University of Oxford), Xiuye Gu (Google Research), and Siyang Li (Google Research)

Quilt-LLaVA: Visual Instruction Tuning by Extracting Localized Narratives from Open-Source Histopathology Videos 13183
Mehmet Saygin Seyfioglu (University of Washington), Wisdom O. Ikezogwo (University of Washington), Fatemeh Ghezloo (University of Washington), Ranjay Krishna (University of Washington), and Linda Shapiro (University of Washington)

Aligning and Prompting Everything All at Once for Universal Visual Perception 13193
Yunhang Shen (Tencent Youtu Lab, China), Chaoyou Fu (Tencent Youtu Lab, China), Peixian Chen (Tencent Youtu Lab, China), Mengdan Zhang (Tencent Youtu Lab, China), Ke Li (Tencent Youtu Lab, China), Xing Sun (Tencent Youtu Lab, China), Yunsheng Wu (Tencent Youtu Lab, China), Shaohui Lin (East China Normal University, China), and Rongrong Ji (Xiamen University, China)

Can I Trust Your Answer? Visually Grounded Video Question Answering 13204
Junbin Xiao (National University of Singapore), Angela Yao (National University of Singapore), Yicong Li (National University of Singapore), and Tat-Seng Chua (National University of Singapore)

Prompt Highlighter: Interactive Control for Multi-Modal LLMs 13215
Yuechen Zhang (The Chinese University of Hong Kong, Hong Kong), Shengju Qian (The Chinese University of Hong Kong, Hong Kong), Bohao Peng (The Chinese University of Hong Kong, Hong Kong), Shu Liu (SmartMore, China), and Jiaya Jia (The Chinese University of Hong Kong, Hong Kong; SmartMore, China)

Language-only Training of Zero-shot Composed Image Retrieval 13225
Geonmo Gu (NAVER Vision), Sanghyuk Chun (NAVER AI Lab), Wonjae Kim (NAVER AI Lab), Yoohoon Kang (NAVER Vision), and Sangdoo Yun (NAVER AI Lab)

MoReVQA: Exploring Modular Reasoning Models for Video Question Answering 13235
Juhong Min (POSTECH), Shyamal Buch (Google), Arsha Nagrani (Google), Minsu Cho (POSTECH), and Cordelia Schmid (Google)

Let's Think Outside the Box: Exploring Leap-of-Thought in Large Language Models with Creative Humor Generation 13246
Shanshan Zhong (Sun Yat-Sen University, China), Zhongzhan Huang (Sun Yat-Sen University, China), Shanghua Gao (Harvard University, USA), Wushao Wen (Sun Yat-Sen University, China), Liang Lin (Sun Yat-Sen University, China), Marinka Zitnik (Harvard University, USA), and Pan Zhou (Singapore Management University, Singapore)

CLOVA: A Closed-LOop Visual Assistant with Tool Usage and Update 13258
Zhi Gao (Peking University and Beijing Institute for General Artificial Intelligence, China), Yuntao Du (Beijing Institute for General Artificial Intelligence, China), Xintong Zhang (Beijing Jiaotong University and Beijing Institute for General Artificial Intelligence, China), Xiaojian Ma (Beijing Institute for General Artificial Intelligence, China), Wenjuan Han (Beijing Jiaotong University, China), Song-Chun Zhu (Peking University, Beijing Institute for General Artificial Intelligence, and Tsinghua University, China), and Qing Li (Beijing Institute for General Artificial Intelligence, China)

Naturally Supervised 3D Visual Grounding with Language-Regularized Concept Learners 13269
Chun Feng (Stanford University, USA), Joy Hsu (Stanford University, USA), Weiyu Liu (Stanford University, USA), and Jiajun Wu (Stanford University, USA)

Synthesize, Diagnose, and Optimize: Towards Fine-Grained Vision-Language Understanding ... 13279
Wujian Peng (Fudan University, China), Sicheng Xie (Fudan University, China), Zuyao You (Fudan University, China), Shiyi Lan (NVIDIA, USA), and Zuxuan Wu (Fudan University, China)

AssistGUI: Task-Oriented PC Graphical User Interface Automation 13289
Difei Gao (National University of Singapore), Lei Ji (Microsoft), Zechen Bai (National University of Singapore), Mingyu Ouyang (National University of Singapore), Peiran Li (National University of Singapore), Dongxing Mao (National University of Singapore), Qinchen Wu (National University of Singapore), Weichen Zhang (National University of Singapore), Peiyi Wang (National University of Singapore), Xiangwu Guo (National University of Singapore), Hengxu Wang (National University of Singapore), Luowei Zhou (Google DeepMind), and Mike Zheng Shou (National University of Singapore)

SEED-Bench: Benchmarking Multimodal Large Language Models 13299
Bohao Li (The Chinese University of HongKong, Shenzhen), Yuying Ge (Tencent AI Lab), Yixiao Ge (Tencent), Guangzhi Wang (National University of Singapore), Rui Wang (Fudan University), Ruimao Zhang (The Chinese University of HongKong, Shenzhen), and Ying Shan (Tencent AI Lab Center of Visual Computing)

Unknown Prompt, the only Lacuna: Unveiling CLIP's Potential for Open Domain Generalization..... 13309
Mainak Singha (Aisin Corporation, Japan), Ankit Jha (IIT Bombay, India), Shirsha Bose (TU Munich, Germany), Ashwin Nair (IISER Thiruvananthapuram, India), Moloud Abdar (Deakin University, Australia), and Biplab Banerjee (IIT Bombay, India)

Panda-70M: Captioning 70M Videos with Multiple Cross-Modality Teachers 13320
Tsai-Shien Chen (University of California, Merced, USA), Aliaksandr Siarohin (Snap Inc., USA), Willi Menapace (Snap Inc., USA), Ekaterina Deyneka (Snap Inc., USA), Hsiang-wei Chao (Snap Inc., USA), Byung Eun Jeon (Snap Inc., USA), Yuwei Fang (Snap Inc., USA), Hsin-Ying Lee (Snap Inc., USA), Jian Ren (Snap Inc., USA), Ming-Hsuan Yang (University of California, Merced, USA), and Sergey Tulyakov (Snap Inc., USA)

Decoupling Static and Hierarchical Motion Perception for Referring Video Segmentation 13332
Shuting He (Nanyang Technological University, Singapore) and Henghui Ding (Fudan University, China)

Causal-CoG: A Causal-Effect Look at Context Generation for Boosting Multi-modal Language Models 13342
Shitian Zhao (East China Normal University, China), Zhuowan Li (Johns Hopkins University, USA), Yadong Lu (East China Normal University, China), Alan Yuille (Johns Hopkins University, USA), and Yan Wang (East China Normal University, China)

Posterior Distillation Sampling 13352
Juil Koo (KAIST), Chanho Park (KAIST), and Minhyuk Sung (KAIST)

Towards More Unified In-context Visual Understanding 13362
Dianmo Sheng (University of Science and Technology of China, China), Dongdong Chen (Microsoft, America), Zhentao Tan (University of Science and Technology of China, China), Qiankun Liu (Beijing Institute of Technology, China), Qi Chu (University of Science and Technology of China, China), Jianmin Bao (Microsoft, China), Tao Gong (University of Science and Technology of China, China), Bin Liu (University of Science and Technology of China, China), Shengwei Xu (Beijing Electronic Science and Technology Institute, China), and Nenghai Yu (University of Science and Technology of China, China)

Mask4Align: Aligned Entity Prompting with Color Masks for Multi-Entity Localization Problems 13373
Haoquan Zhang (South China University of Technology), Ronggang Huang (South China University of Technology), Yi Xie (South China University of Technology), and Huaidong Zhang (South China University of Technology)

SOK-Bench: A Situated Video Reasoning Benchmark with Aligned Open-World Knowledge 13384
Andong Wang (The University of Hong Kong), Bo Wu (MIT-IBM Watson AI Lab), Sunli Chen (Institute for Interdisciplinary Information Sciences, Tsinghua University), Zhenfang Chen (MIT-IBM Watson AI lab), Haotian Guan (The University of Hong Kong), Wei-Ning Lee (The University of Hong Kong), Li Erran Li (AWS AI), and Chuang Gan (UMass Amherst & MIT-IBM Watson AI Lab)

Align and Aggregate: Compositional Reasoning with Video Alignment and Answer Aggregation for Video Question-Answering 13395
Zhaohe Liao (Shanghai Jiao Tong University), Jiangtong Li (Tongji University), Li Niu (Shanghai Jiao Tong University), and Liqing Zhang (Shanghai Jiao Tong University)

Segment and Caption Anything 13405
Xiaoke Huang (Tsinghua University), Jianfeng Wang (Microsoft), Yansong Tang (Tsinghua University), Zheng Zhang (Microsoft), Han Hu (Microsoft), Jiwen Lu (Tsinghua University), Lijuan Wang (Microsoft), and Zicheng Liu (Advanced Micro Devices)

OPERA: Alleviating Hallucination in Multi-Modal Large Language Models via Over-Trust Penalty and Retrospection-Allocation 13418
Qidong Huang (University of Science and Technology of China, China; Shanghai AI Laboratory, China), Xiaoyi Dong (Shanghai AI Laboratory, China; The Chinese University of Hong Kong, China), Pan Zhang (Shanghai AI Laboratory, China), Bin Wang (Shanghai AI Laboratory, China), Conghui He (Shanghai AI Laboratory, China), Jiaqi Wang (Shanghai AI Laboratory, China), Dahua Lin (Shanghai AI Laboratory, China), Weiming Zhang (University of Science and Technology of China, China), and Nenghai Yu (University of Science and Technology of China, China)

Learning by Correction: Efficient Tuning Task for Zero-Shot Generative Vision-Language Reasoning 13428
Rongjie Li (ShanghaiTech University), Yu Wu (ShanghaiTech University), and Xuming He (ShanghaiTech University; Shanghai Engineering Research Center of Intelligent Vision and Imaging)

Revisiting Counterfactual Problems in Referring Expression Comprehension 13438
Zhihan Yu (Beijing University of Posts and Telecommunications, China) and Ruifan Li (Beijing University of Posts and Telecommunications, China)

ScanFormer: Referring Expression Comprehension by Iteratively Scanning 13449
Wei Su (Zhejiang University, China), Peihan Miao (Zhejiang University, China), Huanzhang Dou (Zhejiang University, China), and Xi Li (Zhejiang University, China; Zhejiang-Singapore Innovation and AI Joint Research Lab, China)

See, Say, and Segment: Teaching LMMs to Overcome False Premises 13459
Tsung-Han Wu (University of California, Berkeley), Giscard Biamby (University of California, Berkeley), David Chan (University of California, Berkeley), Lisa Dunlap (University of California, Berkeley), Ritwik Gupta (University of California, Berkeley), Xudong Wang (University of California, Berkeley), Joseph E. Gonzalez (University of California, Berkeley), and Trevor Darrell (University of California, Berkeley)

SignGraph: A Sign Sequence is Worth Graphs of Nodes 13470
Shiwei Gan (Nanjing University, China), Yafeng Yin (Nanjing University, China), Zhiwei Jiang (Nanjing University, China), Hongkai Wen (The University of Warwick, UK), Lei Xie (Nanjing University, China), and Sanglu Lu (Nanjing University, China)

Enhancing Vision-Language Pre-training with Rich Supervisions 13480
Yuan Gao (Stanford University), Kunyu Shi (AWS AI Labs), Pengkai Zhu (AWS AI Labs), Edouard Belval (AWS AI Labs), Oren Nuriel (AWS AI Labs), Srikar Appalaraju (AWS AI Labs), Shabnam Ghadar (AWS AI Labs), Zhuowen Tu (AWS AI Labs), Vijay Mahadevan (AWS AI Labs), and Stefano Soatto (AWS AI Labs)

De-Diffusion Makes Text a Strong Cross-Modal Interface 13492
Chen Wei (Johns Hopkins University), Chenxi Liu (Google DeepMind), Siyuan Qiao (Google DeepMind), Zhishuai Zhang (Google DeepMind), Alan Yuille (Johns Hopkins University), and Jiahui Yu (Google DeepMind)

MA-LMM: Memory-Augmented Large Multimodal Model for Long-Term Video Understanding 13504
Bo He (University of Maryland, College Park), Hengduo Li (Meta), Young Kyun Jang (Meta), Menglin Jia (Meta), Xuefei Cao (Meta), Ashish Shah (Meta), Abhinav Shrivastava (University of Maryland, College Park), and Ser-Nam Lim (University of Central Florida)

Incorporating Geo-Diverse Knowledge into Prompting for Increased Geographical Robustness in Object Recognition 13515
Kyle Buettner (University of Pittsburgh, USA), Sina Malakouti (University of Pittsburgh, USA), Xiang Lorraine Li (University of Pittsburgh, USA), and Adriana Kovashka (University of Pittsburgh, USA)

Retrieval-Augmented Egocentric Video Captioning 13525
Jilan Xu (Fudan University, China), Yifei Huang (Shanghai AI Laboratory, China), Junlin Hou (Fudan University, China), Guo Chen (Nanjing University, China), Yuejie Zhang (Fudan University, China), Rui Feng (Fudan University, China), and Weidi Xie (Shanghai Jiao Tong University, China)

Towards Better Vision-Inspired Vision-Language Models 13537
Yun-Hao Cao (Nanjing University, China), Kaixiang Ji (Ant Group, China), Ziyuan Huang (Ant Group, China), Chuanyang Zheng (Ant Group, China), Jiajia Liu (Ant Group, China), Jian Wang (Ant Group, China), Jingdong Chen (Ant Group, China), and Ming Yang (Ant Group, China)

PIN: Positional Insert Unlocks Object Localisation Abilities in VLMs 13548
Michael Dorkenwald (University of Amsterdam, Netherlands), Nimrod Barazani (University of Amsterdam, Netherlands), Cees G. M. Snoek (University of Amsterdam, Netherlands), and Yuki M. Asano (University of Amsterdam, Netherlands)

Polos: Multimodal Metric Learning from Human Feedback for Image Captioning 13559
Yuiga Wada (Keio University), Kanta Kaneda (Keio University), Daichi Saito (Keio University), and Komei Sugiura (Keio University)

Siamese Learning with Joint Alignment and Regression for Weakly-Supervised Video Paragraph Grounding 13569
Chaolei Tan (Sun Yat-sen University), Jianhuang Lai (Sun Yat-sen University), Wei-Shi Zheng (Sun Yat-sen University), and Jian-Fang Hu (Sun Yat-sen University)

Koala: Key Frame-Conditioned Long Video-LLM 13581
Reuben Tan (Boston University), Ximeng Sun (Boston University), Ping Hu (University of Electronic Science and Technology of China), Jui-hsien Wang (Adobe Research), Hanieh Deilamsalehy (Adobe Research), Bryan A. Plummer (Boston University), Bryan Russell (Adobe Research), and Kate Saenko (Boston University)

Generating Enhanced Negatives for Training Language-Based Object Detectors 13592
Shiyu Zhao (Rutgers University, USA), Long Zhao (Google Research, USA), Vijay Kumar B G (NEC Laboratories America, USA), Yumin Suh (NEC Laboratories America, USA), Dimitris N. Metaxas (Rutgers University, USA), Manmohan Chandraker (UC San Diego, USA), and Samuel Schulter (NEC Laboratories America, USA)

Non-autoregressive Sequence-to-Sequence Vision-Language Models 13603
Kunyu Shi (AWS AI Labs), Qi Dong (AWS AI Labs), Luis Goncalves (AWS AI Labs), Zhuowen Tu (AWS AI Labs), and Stefano Soatto (AWS AI Labs)

Synthesize Step-by-Step: Tools, Templates and LLMs as Data Generators for Reasoning-Based Chart VQA 13613
Zhuowan Li (Johns Hopkin University), Bhavan Jasani (AWS AI Labs), Peng Tang (AWS AI Labs), and Shabnam Ghadar (AWS AI Labs)

Towards Learning a Generalist Model for Embodied Navigation 13624
Duo Zheng (The Chinese University of Hong Kong, China), Shijia Huang (The Chinese University of Hong Kong, China), Lin Zhao (Centre for Perceptual and Interactive Intelligence, China), Yiwu Zhong (The Chinese University of Hong Kong, China), and Liwei Wang (The Chinese University of Hong Kong, China)

"Previously on ..." From Recaps to Story Summarization 13635
Aditya Kumar Singh (International Institute of Information Technology, Hyderabad), Dhruv Srivastava (International Institute of Information Technology, Hyderabad), and Makarand Tapaswi (International Institute of Information Technology, Hyderabad)

MM-Narrator: Narrating Long-form Videos with Multimodal In-Context Learning 13647
Chaoyi Zhang (University of Sydney, Australia), Kevin Lin (Microsoft, USA), Zhengyuan Yang (Microsoft, USA), Jianfeng Wang (Microsoft, USA), Linjie Li (Microsoft, USA), Chung-Ching Lin (Microsoft, USA), Zicheng Liu (AMD, USA), and Lijuan Wang (Microsoft, USA)

BT-Adapter: Video Conversation is Feasible Without Video Instruction Tuning 13658
Ruyang Liu (Peking University), Chen Li (Applied Research Center (ARC), Tencent PCG), Yixiao Ge (Applied Research Center (ARC), Tencent PCG), Thomas H. Li (Peking University), Ying Shan (Applied Research Center (ARC), Tencent PCG), and Ge Li (Peking University)

Holistic Autonomous Driving Understanding by Bird's-Eye-View Injected Multi-Modal Large Models 13668
Xinpeng Ding (HKUST), Jianhua Han (Huawei Noah's Ark Lab), Hang Xu (Huawei Noah's Ark Lab), Xiaodan Liang (Sun Yat-Sen University), Wei Zhang (Huawei Noah's Ark Lab), and Xiaomeng Li (The Hong Kong University of Science and Technology)

Situational Awareness Matters in 3D Vision Language Reasoning 13678
Yunze Man (University of Illinois at Urbana-Champaign, USA), Liang-Yan Gui (University of Illinois at Urbana-Champaign, USA), and Yu-Xiong Wang (University of Illinois at Urbana-Champaign, USA)

SRTube: Video-Language Pre-Training with Action-Centric Video Tube Features and Semantic Role Labeling 13689
Ju-Hee Lee (Ewha Womans University, South Korea) and Je-Won Kang (Ewha Womans University, South Korea)

Chat-UniVi: Unified Visual Representation Empowers Large Language Models with Image and Video Understanding 13700
Peng Jin (Peking University, China), Ryuichi Takanobu (Tsinghua University, China), Wancai Zhang (Nari Technology Co.,Ltd., China), Xiaochun Cao (Sun Yat-sen University, China), and Li Yuan (Peking University, China)

Curriculum Point Prompting for Weakly-Supervised Referring Image Segmentation 13711
Qiyuan Dai (ShanghaiTech University, China) and Sibei Yang (ShanghaiTech University, China)

Tune-An-Ellipse: CLIP Has Potential to Find What You Want 13723
Jinheng Xie (Show Lab, National University of Singapore), Songhe Deng (Shenzhen University), Bing Li (AI Initiative, King Abdullah University of Science and Technology), Haozhe Liu (AI Initiative, King Abdullah University of Science and Technology), Yawen Huang (Jarvis Research Center, Tencent YouTu Lab), Yefeng Zheng (Jarvis Research Center, Tencent YouTu Lab), Jurgen Schmidhuber (AI Initiative, King Abdullah University of Science and Technology), Bernard Ghanem (AI Initiative, King Abdullah University of Science and Technology), Linlin Shen (Shenzhen University), and Mike Zheng Shou (Show Lab, National University of Singapore)

EVCap: Retrieval-Augmented Image Captioning with External Visual-Name Memory for Open-World Comprehension 13733
Jiaxuan Li (The University of Tokyo, Japan), Duc Minh Vo (The University of Tokyo, Japan), Akihiro Sugimoto (National Institute of Informatics, Japan), and Hideki Nakayama (The University of Tokyo, Japan)

Plug-and-Play Diffusion Distillation 13743
Yi-Ting Hsiao (University of Michigan, USA), Siavash Khodadadeh (Adobe Inc, USA), Kevin Duarte (Adobe Inc, USA), Wei-An Lin (Adobe Inc, USA), Hui Qu (Adobe Inc, USA), Mingi Kwon (Yonsei University, South Korea), and Ratheesh Kalarot (Adobe Inc, USA)

Lookahead Exploration with Neural Radiance Representation for Continuous Vision-Language Navigation 13753
Zihan Wang (Institute of Computing Technology, Chinese Academy of Sciences), Xiangyang Li (Institute of Computing Technology, Chinese Academy of Sciences), Jiahao Yang (Institute of Computing Technology, Chinese Academy of Sciences), Yeqi Liu (Institute of Computing Technology, Chinese Academy of Sciences), Junjie Hu (University of Wisconsin, Madison), Ming Jiang (Indiana University), and Shuqiang Jiang (Institute of Computing Technology, Chinese Academy of Sciences)

Low-Rank Approximation for Sparse Attention in Multi-Modal LLMs 13763
Lin Song (Tencent AILab), Yukang Chen (CUHK), Shuai Yang (HKUST (GZ)), Xiaohan Ding (Tencent AILab), Yixiao Ge (Tencent AILab), Ying-Cong Chen (HKUST (GZ)), and Ying Shan (Tencent AILab)

Contrasting Intra-Modal and Ranking Cross-Modal Hard Negatives to Enhance Visio-Linguistic Compositional Understanding 13774
Le Zhang (Mila - Quebec AI Institute, Canada), Rabiul Awal (Mila - Quebec AI Institute), and Aishwarya Agrawal (Mila - Quebec AI Institute, Canada)

Iterated Learning Improves Compositionality in Large Vision-Language Models 13785
Chenhao Zheng (University of Michigan), Jieyu Zhang (University of Washington), Aniruddha Kembhavi (Allen Institute for Artificial Intelligence), and Ranjay Krishna (University of Washington; Allen Institute for Artificial Intelligence)

RegionGPT: Towards Region Understanding Vision Language Model 13796
Qiushan Guo (The University of Hong Kong), Shalini De Mello (Nvidia Research), Hongxu Yin (Nvidia Research), Wonmin Byeon (Nvidia Research), Ka Chun Cheung (Nvidia), Yizhou Yu (The University of Hong Kong), Ping Luo (The University of Hong Kong), and Sifei Liu (Nvidia Research)

RLHF-V: Towards Trustworthy MLLMs via Behavior Alignment from Fine-grained Correctional Human Feedback 13807
Tianyu Yu (Tsinghua University), Yuan Yao (National University of Singapore), Haoye Zhang (Tsinghua University), Taiwen He (Tsinghua University), Yifeng Han (Tsinghua University), Ganqu Cui (Tsinghua University), Jinyi Hu (Tsinghua University), Zhiyuan Liu (Tsinghua University), Hai-Tao Zheng (Shenzhen International Graduate School, Tsinghua University; Pengcheng Laboratory, Shenzhen, China), Maosong Sun (Tsinghua University), and Tat-Seng Chua (National University of Singapore)

Honeybee: Locality-enhanced Projector for Multimodal LLM 13817
Junbum Cha (Kakao Brain, South Korea), Wooyoung Kang (Kakao Brain, South Korea), Jonghwan Mun (Kakao Brain, South Korea), and Byungseok Roh (Kakao Brain, South Korea)

E-GPS: Explainable Geometry Problem Solving via Top-Down Solver and Bottom-Up Generator 13828
Wenjun Wu (Xi'an Jiaotong University, China), Lingling Zhang (Xi'an Jiaotong University, China), Jun Liu (Xi'an Jiaotong University, China), Xi Tang (Xi'an Jiaotong University, China), Yaxian Wang (Xi'an Jiaotong University, China), Shaowei Wang (Xi'an Jiaotong University, China), and Qianying Wang (Lenovo Research, China)

Pink: Unveiling the Power of Referential Comprehension for Multi-modal LLMs 13838
Shiyu Xuan (Peking University, China), Qingpei Guo (Ant Group), Ming Yang (Ant Group), and Shiliang Zhang (Peking University, China)

Any-Shift Prompting for Generalization over Distributions 13849
Zehao Xiao (University of Amsterdam, Netherlands), Jiayi Shen (University of Amsterdam, Netherlands), Mohammad Mahdi Derakhshani (University of Amsterdam, Netherlands), Shengcai Liao (Core42, UAE), and Cees G. M. Snoek (University of Amsterdam, Netherlands)

Question Aware Vision Transformer for Multimodal Reasoning 13861
Roy Ganz (Technion, Israel), Yair Kittenplon (Amazon AWS, Israel), Aviad Aberdam (Amazon AWS, Israel), Elad Ben Avraham (Amazon AWS, Israel), Oren Nuriel (Amazon AWS, Israel), Shai Mazor (Amazon AWS, Israel), and Ron Litman (Amazon AWS, Israel)

Mitigating Object Hallucinations in Large Vision-Language Models through Visual Contrastive Decoding 13872
Sicong Leng (Nanyang Technological University), Hang Zhang (DAMO Academy, Alibaba Group), Guanzheng Chen (DAMO Academy, Alibaba Group), Xin Li (DAMO Academy, Alibaba Group), Shijian Lu (Nanyang Technological University), Chunyan Miao (Nanyang Technological University), and Lidong Bing (DAMO Academy, Alibaba Group)

Text-Image Alignment for Diffusion-Based Perception 13883
Neehar Kondapaneni (California Institute of Technology, USA), Markus Marks (California Institute of Technology, USA), Manuel Knott (ETH Zurich, Swiss Data Science Center, Empa, Switzerland), Rogerio Guimaraes (California Institute of Technology, USA), and Pietro Perona (California Institute of Technology, USA)

Do You Remember? Dense Video Captioning with Cross-Modal Memory Retrieval 13894
Minkuk Kim (Kyung Hee University, Republic of Korea), Hyeon Bae Kim (Kyung Hee University, Republic of Korea), Jinyoung Moon (Electronics and Telecommunications Research Institute (ETRI), Republic of Korea), Jinwoo Choi (Kyung Hee University, Republic of Korea), and Seong Tae Kim (Kyung Hee University, Republic of Korea)

FairDeDup: Detecting and Mitigating Vision-Language Fairness Disparities in Semantic Dataset Deduplication 13905
Eric Slyman (Oregon State University), Stefan Lee (Oregon State University), Scott Cohen (Adobe Research), and Kushal Kafle (Adobe Research)

G^3-LQ: Marrying Hyperbolic Alignment with Explicit Semantic-Geometric Modeling for 3D Visual Grounding 13917
Yuan Wang (Tsinghua University, China; Beijing National Research Center for Information Science and Technology (BNRist), China), Yali Li (Tsinghua University, China; Beijing National Research Center for Information Science and Technology (BNRist), China), and Shengjin Wang (Tsinghua University, China; Beijing National Research Center for Information Science and Technology (BNRist), China)

VideoCon: Robust Video-Language Alignment via Contrast Captions 13927
Hritik Bansal (University of California Los Angeles), Yonatan Bitton (Google Research), Idan Szpektor (Google Research), Kai-Wei Chang (University of California Los Angeles), and Aditya Grover (University of California Los Angeles)

Taming Self-Training for Open-Vocabulary Object Detection 13938
Shiyu Zhao (Rutgers University, USA), Samuel Schulter (NEC Laboratories America, USA), Long Zhao (Google Research, USA), Zhixing Zhang (Rutgers University, USA), Vijay Kumar B G (NEC Laboratories America, USA), Yumin Suh (NEC Laboratories America, USA), Manmohan Chandraker (UC San Diego, USA), and Dimitris N. Metaxas (Rutgers University, USA)

SyncMask: Synchronized Attentional Masking for Fashion-centric Vision-Language Pretraining. 13948
Chull Hwan Song (Dealicious Inc.), Taebaek Hwang (Dealicious Inc.), Jooyoung Yoon (Dealicious Inc.), Shunghyun Choi (Dealicious Inc.), and Yeong Hyeon Gu (Sejong University)

Generative Region-Language Pretraining for Open-Ended Object Detection 13958
Chuang Lin (Monash University), Yi Jiang (ByteDance Inc.), Lizhen Qu (Monash University), Zehuan Yuan (ByteDance Inc.), and Jianfei Cai (Monash University)

CoG-DQA: Chain-of-Guiding Learning with Large Language Models for Diagram Question Answering 13969
Shaowei Wang (Xi'an Jiaotong University, China), Lingling Zhang (Xi'an Jiaotong University, China), Longji Zhu (Xi'an Jiaotong University, China), Tao Qin (Xi'an Jiaotong University, China), Kim-Hui Yap (Nanyang Technological University, Singapore), Xinyu Zhang (Xi'an Jiaotong University, China), and Jun Liu (Xi'an Jiaotong University, China)

Multi-modal Instruction Tuned LLMs with Fine-grained Visual Perception 13980
Junwen He (Dalian University of Technology, China), Yifan Wang (Dalian University of Technology, China), Lijun Wang (Dalian University of Technology, China), Huchuan Lu (Dalian University of Technology, China), Jun-Yan He (DAMO Academy, Alibaba Group, China), Jin-Peng Lan (DAMO Academy, Alibaba Group, China), Bin Luo (DAMO Academy, Alibaba Group, China), and Xuansong Xie (DAMO Academy, Alibaba Group, China)

Generate Subgoal Images before Act: Unlocking the Chain-of-Thought Reasoning in Diffusion Model for Robot Manipulation with Multimodal Prompts 13991
Fei Ni (Tianjin University, China), Jianye Hao (Tianjin University, China), Shiguang Wu (Huawei Noah's Ark Lab, China), Longxin Kou (Tianjin University, China), Jiashun Liu (Tianjin University, China), Yan Zheng (Tianjin University, China), Bin Wang (Huawei Noah's Ark Lab, China), and Yuzheng Zhuang (Huawei Noah's Ark Lab, China)

LoSh: Long-Short Text Joint Prediction Network for Referring Video Object Segmentation 14001
Linfeng Yuan (Tongji University), Miaojing Shi (Tongji University), Zijie Yue (Tongji University), and Qijun Chen (Tongji University)

MICap: A Unified Model for Identity-Aware Movie Descriptions 14011
Haran Raajesh (International Institute of Information Technology, Hyderabad), Naveen Reddy Desanur (International Institute of Information Technology, Hyderabad), Zeeshan Khan (Inria Paris and Département d'informatique de l'ENS, CNRS, PSL Research University), and Makarand Tapaswi (International Institute of Information Technology, Hyderabad)

CapsFusion: Rethinking Image-Text Data at Scale 14022
Qiying Yu (Institute for AI Industry Research (AIR), Tsinghua University), Quan Sun (Beijing Academy of Artificial Intelligence), Xiaosong Zhang (Beijing Academy of Artificial Intelligence), Yufeng Cui (Beijing Academy of Artificial Intelligence), Fan Zhang (Beijing Academy of Artificial Intelligence), Yue Cao (Independent Researcher), Xinlong Wang (Beijing Academy of Artificial Intelligence), and Jingjing Liu (Institute for AI Industry Research (AIR), Tsinghua University)

Visual Fact Checker: Enabling High-Fidelity Detailed Caption Generation 14033
Yunhao Ge (NVIDIA), Xiaohui Zeng (NVIDIA), Jacob Samuel Huffman (NVIDIA), Tsung-Yi Lin (NVIDIA), Ming-Yu Liu (NVIDIA), and Yin Cui (NVIDIA)

VidLA: Video-Language Alignment at Scale 14043
Mamshad Nayeem Rizve (Amazon, USA), Fan Fei (Amazon, USA), Jayakrishnan Unnikrishnan (Amazon, USA), Son Tran (Amazon, USA), Benjamin Z. Yao (Amazon, USA), Belinda Zeng (Amazon, USA), Mubarak Shah (Amazon, USA; University of Central Florida, USA), and Trishul Chilimbi (Amazon, USA)

Viewpoint-Aware Visual Grounding in 3D Scenes 14056
Xiangxi Shi (Oregon State University), Zhonghua Wu (Sensetime Research, Singapore), and Stefan Lee (Oregon State University)

Multi-Modal Proxy Learning Towards Personalized Visual Multiple Clustering 14066
Jiawei Yao (University of Washington, USA), Qi Qian (Alibaba Group, USA), and Juhua Hu (University of Washington, USA)

Jack of All Tasks, Master of Many: Designing General-Purpose Coarse-to-Fine Vision-Language Model 14076
Shraman Pramanick (Johns Hopkins University, USA), Guangxing Han (Meta, USA), Rui Hou (Meta, USA), Sayan Nag (University of Toronto, Canada), Ser-Nam Lim (University of Central Florida, USA), Nicolas Ballas (Meta, Canada), Qifan Wang (Meta, USA), Rama Chellappa (Johns Hopkins University, USA), and Amjad Almahairi (Meta, USA)

LLaMA-Excitor: General Instruction Tuning via Indirect Feature Interaction 14089
Bo Zou (Tsinghua University, China), Chao Yang (Shanghai AI Laboratory, China), Yu Qiao (Shanghai AI Laboratory, China), Chengbin Quan (Tsinghua University, China), and Youjian Zhao (Tsinghua University, China)

MeaCap: Memory-Augmented Zero-shot Image Captioning 14100
Zequn Zeng (Xidian University, China), Yan Xie (Xidian University, China), Hao Zhang (Xidian University, China), Chiyu Chen (Xidian University, China), Bo Chen (Xidian University, China), and Zhengjue Wang (Xidian University, China)

The STVchrono Dataset: Towards Continuous Change Recognition in Time 14111
Yanjun Sun (National Institute of Advanced Industrial Science and Technology (AIST), Keio University), Yue Qiu (National Institute of Advanced Industrial Science and Technology (AIST)), Mariia Khan (Edith Cowan University), Fumiya Matsuzawa (National Institute of Advanced Industrial Science and Technology (AIST)), and Kenji Iwata (National Institute of Advanced Industrial Science and Technology (AIST))

InstaGen: Enhancing Object Detection by Training on Synthetic Dataset 14121
Chengjian Feng (Meituan Inc.), Yujie Zhong (Meituan Inc.), Zequn Jie (Meituan Inc.), Weidi Xie (Shanghai Jiao Tong University), and Lin Ma (Meituan Inc.)

MiKASA: Multi-Key-Anchor & Scene-Aware Transformer for 3D Visual Grounding 14131
Chun-Peng Chang (German Research Center for Artificial Intelligence, Germany), Shaoxiang Wang (German Research Center for Artificial Intelligence, Germany), Alain Pagani (German Research Center for Artificial Intelligence, Germany), and Didier Stricker (German Research Center for Artificial Intelligence, Germany)

Investigating Compositional Challenges in Vision-Language Models for Visual Grounding 14141
Yunan Zeng (Center for Research on Intelligent Perception and Computing (CRIPAC), Institute of Automation, Chinese Academy of Sciences (CASIA), Meituan), Yan Huang (Center for Research on Intelligent Perception and Computing (CRIPAC), Institute of Automation, Chinese Academy of Sciences (CASIA)), Jinjin Zhang (Meituan), Zequn Jie (Meituan), Zhenhua Chai (Meituan), and Liang Wang (Center for Research on Intelligent Perception and Computing (CRIPAC), Institute of Automation, Chinese Academy of Sciences (CASIA))

Masked AutoDecoder is Effective Multi-Task Vision Generalist 14152
Han Qiu (S-Lab Nanyang Technological University, Singapore), Jiaxing Huang (S-Lab Nanyang Technological University, Singapore), Peng Gao (Shanghai Artificial Intelligence Laboratory), Lewei Lu (Sensetime Research), Xiaoqin Zhang (Zhejiang University of Technology), and Shijian Lu (S-Lab Nanyang Technological University, Singapore)

Efficient Test-Time Adaptation of Vision-Language Models 14162
Adilbek Karmanov (Mohamed bin Zayed University of Artificial Intelligence), Dayan Guan (Nanyang Technological University), Shijian Lu (Mohamed bin Zayed University of Artificial Intelligence), Abdulmotaleb El Saddik (Mohamed bin Zayed University of Artificial Intelligence), and Eric Xing (Mohamed bin Zayed University of Artificial Intelligence)

FFF: Fixing Flawed Foundations in Contrastive Pre-Training Results in Very Strong Vision-Language Models 14172
Adrian Bulat (Samsung AI Cambridge), Yassine Ouali (Samsung AI Cambridge), and Georgios Tzimiropoulos (Samsung AI Cambridge; Queen Mary University of London)

Open3DSG: Open-Vocabulary 3D Scene Graphs from Point Clouds with Queryable Objects and Open-Set Relationships 14183
Sebastian Koch (Ulm University / Bosch Center for AI), Narunas Vaskevicius (Bosch Center for AI), Mirco Colosi (Robert Bosch Corporate Research), Pedro Hermosilla (TU Wien), and Timo Ropinski (Ulm University)

Instance-level Expert Knowledge and Aggregate Discriminative Attention for Radiology Report Generation 14194
Shenshen Bu (Sun Yat-sen University, China), Taiji Li (Sun Yat-sen University, China), Yuedong Yang (Sun Yat-sen University, China), and Zhiming Dai (Sun Yat-sen University, China)

Omni-SMoLA: Boosting Generalist Multimodal Models with Soft Mixture of Low-rank Experts . 14205
Jialin Wu (Google Research), Xia Hu (Google Research), Yaqing Wang (Google Deepmind), Bo Pang (Google Research), and Radu Soricut (Google Research)

Building Vision-Language Models on Solid Foundations with Masked Distillation 14216
Sepehr Sameni (University of Bern, Switzerland), Kushal Kafle (Adobe Research, USA), Hao Tan (Adobe Research, USA), and Simon Jenni (Adobe Research, Switzerland)

GROUNDHOG: Grounding Large Language Models to Holistic Segmentation 14227
Yichi Zhang (University of Michigan, USA), Ziqiao Ma (University of Michigan, USA), Xiaofeng Gao (Amazon AGI, USA), Suhaila Shakiah (Amazon AGI, USA), Qiaozi Gao (Amazon AGI, USA), and Joyce Chai (University of Michigan, USA)

DRESS: Instructing Large Vision-Language Models to Align and Interact with Humans via Natural Language Feedback 14239
Yangyi Chen (University of Illinois Urbana-Champaign, USA), Karan Sikka (SRI International, USA), Michael Cogswell (SRI International, USA), Heng Ji (University of Illinois Urbana-Champaign, USA), and Ajay Divakaran (SRI International, USA)

LASO: Language-guided Affordance Segmentation on 3D Object 14251
Yicong Li (National University of Singapore), Na Zhao (Singapore University of Technology and Design), Junbin Xiao (National University of Singapore), Chun Feng (University of Science and Technology of China), Xiang Wang (University of Science and Technology of China), and Tat-seng Chua (National University of Singapore)

Omni-Q: Omni-Directional Scene Understanding for Unsupervised Visual Grounding 14261
Sai Wang (Wuhan University, China), Yutian Lin (Wuhan University, China), and Yu Wu (Wuhan University, China)

VTimeLLM: Empower LLM to Grasp Video Moments 14271
Bin Huang (Tsinghua University, China), Xin Wang (Tsinghua University, China), Hong Chen (Tsinghua University, China), Zihan Song (Tsinghua University, China), and Wenwu Zhu (Tsinghua University, China)

CogAgent: A Visual Language Model for GUI Agents 14281
Wenyi Hong (Tsinghua University), Weihan Wang (Tsinghua University), Qingsong Lv (Zhipu AI), Jiazheng Xu (Tsinghua University), Wenmeng Yu (Zhipu AI), Junhui Ji (Zhipu AI), Yan Wang (Zhipu AI), Zihan Wang (Tsinghua University), Yuxiao Dong (Tsinghua University), Ming Ding (Zhipu AI), and Jie Tang (Tsinghua University)

EgoThink: Evaluating First-Person Perspective Thinking Capability of Vision-Language Models 14291
Sijie Cheng (Tsinghua University, China), Zhicheng Guo (Tsinghua University, China), Jingwen Wu (University of Toronto, Canada), Kechen Fang (Tsinghua University, China), Peng Li (Tsinghua University, China), Huaping Liu (Tsinghua University, China), and Yang Liu (Tsinghua University, China)

Multi-Modal Hallucination Control by Visual Information Grounding 14303
Alessandro Favero (EPFL), Luca Zancato (AWS AI Labs), Matthew Trager (AWS AI Labs), Siddharth Choudhary (AWS AI Labs), Pramuditha Perera (AWS AI Labs), Alessandro Achille (AWS AI Labs), Ashwin Swaminathan (AWS AI Labs), and Stefano Soatto (AWS AI Labs)

TimeChat: A Time-sensitive Multimodal Large Language Model for Long Video Understanding 14313
Shuhuai Ren (Peking University, China), Linli Yao (Peking University, China), Shicheng Li (Peking University, China), Xu Sun (Peking University, China), and Lu Hou (Huawei, China)

AHIVE: Anatomy-aware Hierarchical Vision Encoding for Interactive Radiology Report Retrieval 14324
Sixing Yan (Hong Kong Baptist University, Hong Kong SAR, China), William K. Cheung (Hong Kong Baptist University, Hong Kong SAR, China), Ivor W. Tsang (Agency for Science, Technology and Research, Singapore), Keith Chiu (Queen Elizabeth and Kwong Wah Hospitals, Hong Kong SAR, China), Terence M. Tong (Tuen Mun Hospital, Hong Kong SAR, China), Ka Chun Cheung (NVIDIA AI Technology Center, NVIDIA Corporation), and Simon See (NVIDIA AI Technology Center, NVIDIA Corporation)

Do Vision and Language Encoders Represent the World Similarly? 14334
Mayug Maniparambil (ML Labs, Dublin City University), Raiymbek Akshulakov (University of California, Berkeley), Yasser Abdelaziz Dahou Djilali (Technological Innovation Institute), Mohamed El Amine Seddik (Technological Innovation Institute), Sanath Narayan (Technological Innovation Institute), Karttikeya Mangalam (University of California, Berkeley), and Noel E. O'Connor (ML Labs, Dublin City University; Technological Innovation Institute)

Self-Training Large Language Models for Improved Visual Program Synthesis With Visual Reinforcement 14344
Zaid Khan (Northeastern University), Vijay Kumar BG (NEC Laboratories America), Samuel Schulter (NEC Laboratories America), Yun Fu (Northeastern University), and Manmohan Chandraker (NEC Laboratories America, UC San Diego)

Composing Object Relations and Attributes for Image-Text Matching 14354
Khoi Pham (University of Maryland, College Park), Chuong Huynh (University of Maryland, College Park), Ser-Nam Lim (University of Central Florida), and Abhinav Shrivastava (University of Maryland, College Park)

Zero-shot Referring Expression Comprehension via Structural Similarity Between Images and Captions 14364
Zeyu Han (Northeastern University, USA), Fangrui Zhu (Northeastern University, USA), Qianru Lao (Harvard University, USA), and Huaizu Jiang (Northeastern University, USA)

HallusionBench: An Advanced Diagnostic Suite for Entangled Language Hallucination and Visual Illusion in Large Vision-Language Models 14375
Tianrui Guan (University of Maryland, USA), Fuxiao Liu (University of Maryland, USA), Xiyang Wu (University of Maryland, USA), Ruiqi Xian (University of Maryland, USA), Zongxia Li (University of Maryland, USA), Xiaoyu Liu (University of Maryland, USA), Xijun Wang (University of Maryland, USA), Lichang Chen (University of Maryland, USA), Furong Huang (University of Maryland, USA), Yaser Yacoob (University of Maryland, USA), Dinesh Manocha (University of Maryland, USA), and Tianyi Zhou (University of Maryland, USA)

A Simple Recipe for Contrastively Pre-training Video-First Encoders Beyond 16 Frames 14386
Pinelopi Papalampidi (Google DeepMind, UK), Skanda Koppula (Google DeepMind, UK), Shreya Pathak (Google DeepMind, UK), Justin Chiu (Google DeepMind, UK), Joe Heyward (Google DeepMind, UK), Viorica Patraucean (Google DeepMind, UK), Jiajun Shen (Google DeepMind, UK), Antoine Miech (Google DeepMind, UK), Andrew Zisserman (Google DeepMind, UK), and Aida Nematzdeh (Google DeepMind, UK)

Generative Multimodal Models are In-Context Learners 14398
Quan Sun (Beijing Academy of Artificial Intelligence), Yufeng Cui (Beijing Academy of Artificial Intelligence), Xiaosong Zhang (Beijing Academy of Artificial Intelligence), Fan Zhang (Beijing Academy of Artificial Intelligence), Qiying Yu (Institute for AI Industry Research (AIR), Tsinghua University), Yueze Wang (Beijing Academy of Artificial Intelligence), Yongming Rao (Beijing Academy of Artificial Intelligence), Jingjing Liu (Institute for AI Industry Research (AIR), Tsinghua University), Tiejun Huang (Beijing Academy of Artificial Intelligence), and Xinlong Wang (Beijing Academy of Artificial Intelligence)

A Vision Check-up for Language Models 14410
Pratyusha Sharma (MIT), Tamar Rott Shaham (MIT), Manel Baradad (MIT), Stephanie Fu (UC Berkeley), Adrian Rodriguez-Munoz (MIT), Shivam Duggal (MIT), Phillip Isola (MIT), and Antonio Torralba (MIT)

Compositional Chain-of-Thought Prompting for Large Multimodal Models 14420
Chancharik Mitra (University of California, Berkeley), Brandon Huang (University of California, Berkeley), Trevor Darrell (University of California, Berkeley), and Roei Herzig (University of California, Berkeley)

On Scaling Up a Multilingual Vision and Language Model 14432
Xi Chen (Google, USA), Josip Djolonga (Google, USA), Piotr Padlewski (Google, USA), Basil Mustafa (Google, USA), Soravit Changpinyo (Google, USA), Jialin Wu (Google, USA), Carlos Riquelme Ruiz (Google, USA), Sebastian Goodman (Google, USA), Xiao Wang (Google, USA), Yi Tay (Google, USA), Siamak Shakeri (Google, USA), Mostafa Dehghani (Google, USA), Daniel Salz (Google, USA), Mario Lucic (Google, USA), Michael Tschannen (Google, USA), Arsha Nagrani (Google, USA), Hexiang Hu (Google, USA), Mandar Joshi (Google, USA), Bo Pang (Google, USA), Ceslee Montgomery (Google, USA), Paulina Pietrzyk (Google, USA), Marvin Ritter (Google, USA), AJ Piergiovanni (Google, USA), Matthias Minderer (Google, USA), Filip Pavetic (Google, USA), Austin Waters (Google, USA), Gang Li (Google, USA), Ibrahim Alabdulmohsin (Google, USA), Lucas Beyer (Google, USA), Julien Amelot (Google, USA), Kenton Lee (Google, USA), Andreas Peter Steiner (Google, USA), Yang Li (Google, USA), Daniel Keysers (Google, USA), Anurag Arnab (Google, USA), Yuanzhong Xu (Google, USA), Keran Rong (Google, USA), Alexander Kolesnikov (Google, USA), Mojtaba Seyedhosseini (Google, USA), Anelia Angelova (Google, USA), Xiaohua Zhai (Google, USA), Neil Houlsby (Google, USA), and Radu Soricut (Google, USA)

Dual-View Visual Contextualization for Web Navigation 14445
Jihyung Kil (The Ohio State University, USA), Chan Hee Song (The Ohio State University, USA), Boyuan Zheng (The Ohio State University, USA), Xiang Deng (The Ohio State University, USA), Yu Su (The Ohio State University, USA), and Wei-Lun Chao (The Ohio State University, USA)

SpatialVLM: Endowing Vision-Language Models with Spatial Reasoning Capabilities 14455
Boyuan Chen (MIT, USA), Zhuo Xu (Google Deepmind, USA), Sean Kirmani (Google Deepmind, USA), Brain Ichter (Google Deepmind, USA), Dorsa Sadigh (Stanford University, USA), Leonidas Guibas (Stanford University, USA), and Fei Xia (Google Deepmind, USA)

Beyond Seen Primitive Concepts and Attribute-Object Compositional Learning 14466
Nirat Saini (University Of Maryland, College Park), Khoi Pham (University Of Maryland, College Park), and Abhinav Shrivastava (University Of Maryland, College Park)

Orals 4A Autonomous Navigation and Egocentric Vision

SAFDNet: A Simple and Effective Network for Fully Sparse 3D Object Detection 14477
Gang Zhang (Tsinghua University), Junnan Chen (Huazhong University of Science and Technology), Guohuan Gao (Beijing Institute of Technology), Jianmin Li (Tsinghua University), Si Liu (Beihang University), and Xiaolin Hu (Tsinghua University)

UnO: Unsupervised Occupancy Fields for Perception and Forecasting 14487
Ben Agro (Waabi), Quinlan Sykora (Waabi), Sergio Casas (Waabi), Thomas Gilles (Waabi), and Raquel Urtasun (Waabi)

EgoGen: An Egocentric Synthetic Data Generator 14497
Gen Li (ETH Zurich, Switzerland), Kaifeng Zhao (ETH Zurich, Switzerland), Siwei Zhang (ETH Zurich, Switzerland), Xiaozhong Lyu (ETH Zurich, Switzerland), Mihai Dusmanu (Microsoft, Switzerland), Yan Zhang (ETH Zurich, Switzerland), Marc Pollefeys (ETH Zurich & Microsoft, Switzerland), and Siyu Tang (ETH Zurich, Switzerland)

Learning to Segment Referred Objects from Narrated Egocentric Videos 14510
Yuhan Shen (Northeastern University), Huiyu Wang (FAIR, Meta), Xitong Yang (FAIR, Meta), Matt Feiszli (FAIR, Meta), Ehsan Elhamifar (Northeastern University), Lorenzo Torresani (FAIR, Meta), and Effrosyni Mavroudi (FAIR, Meta)

Producing and Leveraging Online Map Uncertainty in Trajectory Prediction 14521
Xunjiang Gu (University of Toronto, Canada), Guanyu Song (University of Toronto, Canada), Igor Gilitschenski (University of Toronto, Canada), Marco Pavone (NVIDIA Research, USA), and Boris Ivanovic (NVIDIA Research, USA)

Orals 4B 3D Vision

SceneFun3D: Fine-Grained Functionality and Affordance Understanding in 3D Scenes 14531
Alexandros Delitzas (ETH Zurich, Switzerland), Ayca Takmaz (ETH Zurich, Switzerland), Federico Tombari (Google, Switzerland; TUM, Germany), Robert Sumner (ETH Zurich, Switzerland), Marc Pollefeys (ETH Zurich, Switzerland; Microsoft, Switzerland), and Francis Engelmann (ETH Zurich, Switzerland; Google, Switzerland)

SpiderMatch: 3D Shape Matching with Global Optimality and Geometric Consistency 14543
Paul Roetzer (University of Bonn) and Florian Bernard (University of Bonn)

PaSCo: Urban 3D Panoptic Scene Completion with Uncertainty Awareness 14554
Anh-Quan Cao (Inria, France), Angela Dai (Technical University of Munich, Germany), and Raoul de Charette (Inria, France)

PlatoNeRF: 3D Reconstruction in Plato's Cave via Single-View Two-Bounce Lidar 14565
Tzofi Klinghoffer (Massachusetts Institute of Technology), Xiaoyu Xiang (Meta), Siddharth Somasundaram (Massachusetts Institute of Technology), Yuchen Fan (Meta), Christian Richardt (Meta), Ramesh Raskar (Massachusetts Institute of Technology), and Rakesh Ranjan (Meta)

A Subspace-Constrained Tyler's Estimator and its Applications to Structure from Motion 14575
Feng Yu (University of Minnesota Twin Cities, USA), Teng Zhang (University of Central Florida, USA), and Gilad Lerman (University of Minnesota Twin Cities, USA)

Orals 4C Action and Motion

Modeling Multimodal Social Interactions: New Challenges and Baselines with Densely Aligned Representations 14585
Sangmin Lee (University of Illinois Urbana-Champaign, USA), Bolin Lai (Georgia Institute of Technology, USA), Fiona Ryan (Georgia Institute of Technology, USA), Bikram Boote (University of Illinois Urbana-Champaign, USA), and James M. Rehg (University of Illinois Urbana-Champaign, USA)

An N-Point Linear Solver for Line and Motion Estimation with Event Cameras 14596
Ling Gao (ShanghaiTech University, China), Daniel Gehrig (University of Zurich, Switzerland), Hang Su (ShanghaiTech University, China), Davide Scaramuzza (University of Zurich, Switzerland), and Laurent Kneip (ShanghaiTech University, China)

RoHM: Robust Human Motion Reconstruction via Diffusion 14606
Siwei Zhang (ETH Zurich, Switzerland), Bharat Lal Bhatnagar (Meta Reality Labs Research, Switzerland), Yuanlu Xu (Meta Reality Labs Research, USA), Alexander Winkler (Meta Reality Labs Research, USA), Petr Kadlecek (Meta Reality Labs Research, Switzerland), Siyu Tang (ETH Zurich, Switzerland), and Federica Bogo (Meta Reality Labs Research, Switzerland)

Temporally Consistent Unbalanced Optimal Transport for Unsupervised Action Segmentation . 14618
Ming Xu (The Australian National University) and Stephen Gould (The Australian National University)

FineParser: A Fine-grained Spatio-temporal Action Parser for Human-centric Action Quality Assessment 14628
Jinglin Xu (University of Science and Technology Beijing, China), Sibo Yin (Peking University, China), Guohao Zhao (Peking University, China), Zishuo Wang (Peking University, China), and Yuxin Peng (Peking University, China)

Poster Session 4

Self-Supervised Class-Agnostic Motion Prediction with Spatial and Temporal Consistency Regularizations 14638
Kewei Wang (Huazhong University of Science and Technology, China), Yizheng Wu (Huazhong University of Science and Technology, China), Jun Cen (Nanyang Technological University, Singapore), Zhiyu Pan (Huazhong University of Science and Technology, China), Xingyi Li (Huazhong University of Science and Technology, China), Zhe Wang (Sensetime Group Limited, China), Zhiguo Cao (Huazhong University of Science and Technology, China), and Guosheng Lin (Nanyang Technological University, Singapore)

Multi-Space Alignments Towards Universal LiDAR Segmentation 14648
Youquan Liu (ShanghaiTech University), Lingdong Kong (National University of Singapore), Xiaoyang Wu (University of Hong Kong), Runnan Chen (University of Hong Kong), Xin Li (East China Normal University), Liang Pan (Shanghai AI Laboratory), Ziwei Liu (Nanyang Technological University, Singapore), and Yuexin Ma (ShanghaiTech University)

Generalized Predictive Model for Autonomous Driving 14662
Jiazhi Yang (Shanghai AI Lab, China), Shenyuan Gao (Shanghai AI Lab, China), Yihang Qiu (Shanghai AI Lab, China), Li Chen (Shanghai AI Lab, China), Tianyu Li (Shanghai AI Lab, China), Bo Dai (Shanghai AI Lab, China), Kashyap Chitta (University of Tübingen, Germany), Penghao Wu (Shanghai AI Lab, China), Jia Zeng (Shanghai AI Lab, China), Ping Luo (University of Hong Kong, China), Jun Zhang (Hong Kong University of Science and Technology, China), Andreas Geiger (University of Tübingen, Germany), Yu Qiao (Shanghai AI Lab, China), and Hongyang Li (Shanghai AI Lab, China)

Visual Point Cloud Forecasting enables Scalable Autonomous Driving 14673
Zetong Yang (Shanghai AI Lab, China), Li Chen (Shanghai AI Lab), Yanan Sun (Shanghai AI Lab), and Hongyang Li (Shanghai AI Lab)

SeMoLi: What Moves Together Belongs Together 14685
Jenny Seidenschwarz (Technical University of Munich), Aljosa Osep (NVIDIA), Francesco Ferroni (NVIDIA), Simon Lucey (University of Adelaide), and Laura Leal-Taixe (NVIDIA)

AIDE: An Automatic Data Engine for Object Detection in Autonomous Driving 14695
Mingfu Liang (Northwestern University, USA), Jong-Chyi Su (NEC Laboratories America, USA), Samuel Schulter (NEC Laboratories America, USA), Sparsh Garg (NEC Laboratories America, USA), Shiyu Zhao (Rutgers University, USA), Ying Wu (Northwestern University, USA), and Manmohan Chandraker (NEC Laboratories America, USA; UC San Diego, USA)

Dynamic Adapter Meets Prompt Tuning: Parameter-Efficient Transfer Learning for Point Cloud Analysis 14707
Xin Zhou (Huazhong University of Science and Technology), Dingkang Liang (Huazhong University of Science and Technology), Wei Xu (Huazhong University of Science and Technology), Xingkui Zhu (Huazhong University of Science and Technology), Yihan Xu (Huazhong University of Science and Technology), Zhikang Zou (Huazhong University of Science and Technology), and Xiang Bai (Huazhong University of Science and Technology)

BEVSpread: Spread Voxel Pooling for Bird's-Eye-View Representation in Vision-based Roadside 3D Object Detection 14718
Wenjie Wang (Zhejiang University, China), Yehao Lu (Zhejiang University, China), Guangcong Zheng (Zhejiang University, China), Shuigen Zhan (Zhejiang University, China), Xiaoqing Ye (Baidu, China), Zichang Tan (Baidu, China), Jingdong Wang (Baidu, China), Gaoang Wang (Zhejiang University, China), and Xi Li (Zhejiang University, China; Zhejiang – Singapore Innovation and AI Joint Research Lab, China;)

DualAD: Disentangling the Dynamic and Static World for End-to-End Driving 14728
Simon Doll (Mercedes-Benz AG, Germany; University of Tübingen, Germany), Niklas Hanselmann (Mercedes-Benz AG, Germany; University of Tübingen, Germany), Lukas Schneider (Mercedes-Benz AG, Germany), Richard Schulz (Mercedes-Benz AG, Germany), Marius Cordts (Mercedes-Benz AG, Germany), Markus Enzweiler (Esslingen University of Applied Sciences, Germany), and Hendrik P. A. Lensch (University of Tübingen, Germany)

Towards Realistic Scene Generation with LiDAR Diffusion Models 14738
Haoxi Ran (Carnegie Mellon University, USA), Vitor Guizilini (Toyota Research Institute, USA), and Yue Wang (University of Southern California, USA)

Driving into the Future: Multiview Visual Forecasting and Planning with World Model for Autonomous Driving 14749
Yuqi Wang (Institute of Automation, Chinese Academy of Sciences, China), Jiawei He (Institute of Automation, Chinese Academy of Sciences, China), Lue Fan (Institute of Automation, Chinese Academy of Sciences, China), Hongxin Li (Institute of Automation, Chinese Academy of Sciences, China), Yuntao Chen (Centre for Artificial Intelligence and Robotics (HKISI_CAS), China), and Zhaoxiang Zhang (Institute of Automation, Chinese Academy of Sciences, China)

VLP: Vision Language Planning for Autonomous Driving 14760
Chenbin Pan (Syracuse University), Burhaneddin Yaman (Bosch Research North America & Bosch Center for Artificial Intelligence (BCAI)), Tommaso Nesti (Bosch Research North America & Bosch Center for Artificial Intelligence (BCAI)), Abhirup Mallik (Bosch Research North America & Bosch Center for Artificial Intelligence (BCAI)), Alessandro G Allievi (Bosch Research North America & Bosch Center for Artificial Intelligence (BCAI)), Senem Velipasalar (Syracuse University), and Liu Ren (Bosch Research North America & Bosch Center for Artificial Intelligence (BCAI))

Scaling Diffusion Models to Real-World 3D LiDAR Scene Completion 14770
Lucas Nunes (University of Bonn), Rodrigo Marcuzzi (University of Bonn), Benedikt Mersch (University of Bonn), Jens Behley (University of Bonn), and Cyrill Stachniss (University of Bonn)

UniMix: Towards Domain Adaptive and Generalizable LiDAR Semantic Segmentation in Adverse Weather 14781
Haimei Zhao (The University of Sydney), Jing Zhang (The University of Sydney), Zhuo Chen (Tsinghua University), Shanshan Zhao (The University of Sydney), and Dacheng Tao (Nanyang Technological University)

Not All Voxels Are Equal: Hardness-Aware Semantic Scene Completion with Self-Distillation ... 14792
Song Wang (Zhejiang University), Jiawei Yu (Zhejiang University), Wentong Li (Zhejiang University), Wenyu Liu (Zhejiang University), Xiaolu Liu (Zhejiang University), Junbo Chen (Udeer.ai), and Jianke Zhu (Zhejiang University)

OOSTraj: Out-of-Sight Trajectory Prediction With Vision-Positioning Denoising 14802
Haichao Zhang (Northeastern University), Yi Xu (Northeastern University), Hongsheng Lu (Toyota Motor North America), Takayuki Shimizu (Toyota Motor North America), and Yun Fu (Northeastern University)

MGMap: Mask-Guided Learning for Online Vectorized HD Map Construction 14812
Xiaolu Liu (Zhejiang University, China), Song Wang (Zhejiang University, China), Wentong Li (Zhejiang University, China), Ruizi Yang (Zhejiang University, China), Junbo Chen (Udeer.ai, China), and Jianke Zhu (Zhejiang University, China)

Density-Adaptive Model Based on Motif Matrix for Multi-Agent Trajectory Prediction 14822
Di Wen (Sun Yat-sen University, China; PengCheng Laboratory, China), Haoran Xu (Sun Yat-sen University, China; PengCheng Laboratory, China), Zhaocheng He (Sun Yat-sen University, China; PengCheng Laboratory, China; Guangdong Provincial Key Laboratory of Intelligent Transportation System, China), Zhe Wu (PengCheng Laboratory, China), Guang Tan (Sun Yat-sen University, China; PengCheng Laboratory, China), and Peixi Peng (PengCheng Laboratory, China; Peking University, China)

StreamingFlow: Streaming Occupancy Forecasting with Asynchronous Multi-modal Data Streams via Neural Ordinary Differential Equation 14833
Yining Shi (Tsinghua University), Kun Jiang (Tsinghua University), Ke Wang (KargoBot, Inc), Jiusi Li (Tsinghua University), Yunlong Wang (Tsinghua University), Mengmeng Yang (Tsinghua University), and Diange Yang (Tsinghua University)

View From Above: Orthogonal-View aware Cross-view Localization 14843
Shan Wang (Data61, CSIRO; Australian National University), Chuong Nguyen (Data61, CSIRO), Jiawei Liu (Australian National University), Yanhao Zhang (Australian National University), Sundaram Muthu (Data61, CSIRO), Fahira Afzal Maken (Data61, CSIRO), Kaihao Zhang (Australian National University), and Hongdong Li (Australian National University)

Improving Distant 3D Object Detection Using 2D Box Supervision 14853
Zetong Yang (Shanghai AI Lab), Zhiding Yu (Nvidia), Chris Choy (Nvidia), Renhao Wang (UC Berkeley), Anima Anandkumar (Caltech), and Jose M. Alvarez (Nvidia)

Is Ego Status All You Need for Open-Loop End-to-End Autonomous Driving? 14864
Zhiqi Li (Nanjing University, China), Zhiding Yu (NVIDIA, USA), Shiyi Lan (NVIDIA, USA), Jiahan Li (Nanjing University, China), Jan Kautz (NVIDIA, USA), Tong Lu (Nanjing University, China), and Jose M. Alvarez (NVIDIA, USA)

CaDeT: a Causal Disentanglement Approach for Robust Trajectory Prediction in Autonomous Driving 14874
Mozhgan Pourkeshavarz (Noah's Ark Laboratory, Huawei, Canada), Junrui Zhang (Noah's Ark Laboratory, Huawei, Canada), and Amir Rasouli (Noah's Ark Laboratory, Huawei, Canada)

Adversarial Backdoor Attack by Naturalistic Data Poisoning on Trajectory Prediction in Autonomous Driving 14885
Mozhgan Pourkeshavarz (Noah's Ark Laboratory, Huawei, Canada), Mohammad Sabokrou (Okinawa Institute of Science and Technology, Japan), and Amir Rasouli (Noah's Ark Laboratory, Huawei, Canada)

NeuRAD: Neural Rendering for Autonomous Driving 14895
Adam Tonderski (Lund University / Zenseact), Carl Lindström (Chalmers University of Technology / Zenseact), Georg Hess (Chalmers University of Technology / Zenseact), William Ljungbergh (Linköping University / Zenseact), Lennart Svensson (Chalmers University of Technology), and Christoffer Petersson (Zenseact)

IS-Fusion: Instance-Scene Collaborative Fusion for Multimodal 3D Object Detection 14905
Junbo Yin (Beijing Institute of Technology), Jianbing Shen (University of Macau), Runnan Chen (The University of Hong Kong), Wei Li (Inceptio), Ruigang Yang (Inceptio), Pascal Frossard (EPFL), and Wenguan Wang (Zhejiang University)

LSK3DNet: Towards Effective and Efficient 3D Perception with Large Sparse Kernels 14916
Tuo Feng (University of Technology Sydney, Australia), Wenguan Wang (Zhejiang University, China), Fan Ma (Zhejiang University, China), and Yi Yang (Zhejiang University, China)

RCBEVDet: Radar-camera Fusion in Bird's Eye View for 3D Object Detection 14928
Zhiwei Lin (Peking University, China), Zhe Liu (University of Electronic Science and Technology of China, China), Zhongyu Xia (Peking University, China), Xinhao Wang (Peking University, China), Yongtao Wang (Peking University, China), Shengxiang Qi (Chongqing Changan Automobile Co., Ltd., China), Yang Dong (Chongqing Changan Automobile Co., Ltd., China), Nan Dong (Chongqing Changan Automobile Co., Ltd., China), Le Zhang (University of Electronic Science and Technology of China, China), and Ce Zhu (University of Electronic Science and Technology of China, China)

PTT: Point-Trajectory Transformer for Efficient Temporal 3D Object Detection 14938
Kuan-Chih Huang (UC Merced, USA), Weijie Lyu (UC Merced, USA), Ming-Hsuan Yang (UC Merced, USA), and Yi-Hsuan Tsai (Google, USA)

Driving Everywhere with Large Language Model Policy Adaptation 14948
Boyi Li (NVIDIA), Yue Wang (University of Southern California / NVIDIA), Jiageng Mao (University of Southern California), Boris Ivanovic (NVIDIA), Sushant Veer (NVIDIA), Karen Leung (NVIDIA / University of Washington), and Marco Pavone (NVIDIA / Stanford University)

Text2Loc: 3D Point Cloud Localization from Natural Language 14958
Yan Xia (Technical University of Munich, Germany; Munich Center for Machine Learning, Germany), Letian Shi (Technical University of Munich, Germany), Zifeng Ding (LMU Munich, Germany), Joao F. Henriques (University of Oxford, UK), and Daniel Cremers (Technical University of Munich, Germany; Munich Center for Machine Learning, Germany)

Commonsense Prototype for Outdoor Unsupervised 3D Object Detection 14968
Hai Wu (Fujian Key Laboratory of Sensing and Computing for Smart Cities, Xiamen University), Shijia Zhao (Fujian Key Laboratory of Sensing and Computing for Smart Cities, Xiamen University), Xun Huang (Fujian Key Laboratory of Sensing and Computing for Smart Cities, Xiamen University), Chenglu Wen (Fujian Key Laboratory of Sensing and Computing for Smart Cities, Xiamen University), Xin Li (Section of Visual Computing and Interactive Media, Texas A&M University), and Cheng Wang (Fujian Key Laboratory of Sensing and Computing for Smart Cities, Xiamen University)

A-Teacher: Asymmetric Network for 3D Semi-Supervised Object Detection 14978
Hanshi Wang (State Key Laboratory of Multimodal Artifcial Intelligence Systems (MAIS), CASIA; School of Artifcial Intelligence, University of Chinese Academy of Sciences), Zhipeng Zhang (KargoBot), Jin Gao (State Key Laboratory of Multimodal Artifcial Intelligence Systems (MAIS), CASIA; School of Artifcial Intelligence, University of Chinese Academy of Sciences), and Weiming Hu (State Key Laboratory of Multimodal Artifcial Intelligence Systems (MAIS), CASIA; School of Artifcial Intelligence, University of Chinese Academy of Sciences; School of Information Science and Technology, ShanghaiTech University)

MoST: Multi-Modality Scene Tokenization for Motion Prediction 14988
Norman Mu (Waymo LLC), Jingwei Ji (Waymo LLC), Zhenpei Yang (Waymo LLC), Nate Harada (Waymo LLC), Haotian Tang (Waymo LLC), Kan Chen (Waymo LLC), Charles R. Qi (Waymo LLC), Runzhou Ge (Waymo LLC), Kratarth Goel (Waymo LLC), Zoey Yang (Waymo LLC), Scott Ettinger (Waymo LLC), Rami Al-Rfou (Waymo LLC), Dragomir Anguelov (Waymo LLC), and Yin Zhou (Waymo LLC)

Feedback-Guided Autonomous Driving 15000
Jimuyang Zhang (Boston University), Zanming Huang (Boston University), Arijit Ray (Boston University), and Eshed Ohn-Bar (Boston University)

Bootstrapping Autonomous Driving Radars with Self-Supervised Learning 15012
Yiduo Hao (University of Cambridge, UK), Sohrab Madani (UIUC, USA), Junfeng Guan (EPFL, Switzerland), Mohammed Alloulah (RadarEye, UK), Saurabh Gupta (UIUC, USA), and Haitham Hassanieh (EPFL, Switzerland)

SIRA: Scalable Inter-frame Relation and Association for Radar Perception 15024
Ryoma Yataka (Mitsubishi Electric Research Laboratories, USA), Pu Wang (Mitsubishi Electric Research Laboratories, USA), Petros Boufounos (Mitsubishi Electric Research Laboratories, USA), and Ryuhei Takahashi (Mitsubishi Electric Corporation, Japan)

SparseOcc: Rethinking Sparse Latent Representation for Vision-Based Semantic Occupancy Prediction 15035
Pin Tang (Shanghai Jiao Tong University, China), Zhongdao Wang (Huawei, China), Guoqing Wang (Shanghai Jiao Tong University, China), Jilai Zheng (Shanghai Jiao Tong University, China), Xiangxuan Ren (Shanghai Jiao Tong University, China), Bailan Feng (Huawei, China), and Chao Ma (Shanghai Jiao Tong University, China)

DiffLoc: Diffusion Model for Outdoor LiDAR Localization 15045
Wen Li (Xiamen University, China), Yuyang Yang (Xiamen University, China), Shangshu Yu (Nanyang Technological University, Singapore), Guosheng Hu (Oosto, U.K), Chenglu Wen (Xiamen University, China), Ming Cheng (Xiamen University, China), and Cheng Wang (Xiamen University, China)

Weak-to-Strong 3D Object Detection with X-Ray Distillation 15055
Alexander Gambashidze (Artificial Intelligence Research Institute, Russia), Aleksandr Dadukin (National Research University Higher School of Economics, Russia), Maxim Golyadkin (Artificial Intelligence Research Institute, Russia), Maria Razzhivina (National Research University Higher School of Economics, Russia), and Ilya Makarov (ISP RAS, Russia)

T4P: Test-Time Training of Trajectory Prediction via Masked Autoencoder and Actor-specific Token Memory 15065
Daehee Park (KAIST), Jaeseok Jeong (KAIST), Sung-Hoon Yoon (KAIST), Jaewoo Jeong (KAIST), and Kuk-Jin Yoon (KAIST)

Editable Scene Simulation for Autonomous Driving via Collaborative LLM-Agents 15077
Yuxi Wei (Shanghai Jiao Tong University, China), Zi Wang (Carnegie Mellon University, USA), Yifan Lu (Shanghai Jiao Tong University, China), Chenxin Xu (Shanghai Jiao Tong University, China), Changxing Liu (Shanghai Jiao Tong University, China), Hao Zhao (Tsinghua University, China), Siheng Chen (Shanghai Jiao Tong University, China; Shanghai AI Laboratory, China), and Yanfeng Wang (Shanghai Jiao Tong University, China; Shanghai AI Laboratory, China)

Uncertainty-Guided Never-Ending Learning to Drive 15088
Lei Lai (Boston University, USA), Eshed Ohn-Bar (Boston University, USA), Sanjay Arora (Red Hat, USA), and John Seon Keun Yi (Boston University, USA)

On the Road to Portability: Compressing End-to-End Motion Planner for Autonomous Driving . 15099
Kaituo Feng (Beijing Institute of Technology, China), Changsheng Li (Beijing Institute of Technology, China), Dongchun Ren (ALLRIDE.AI, China), Ye Yuan (Beijing Institute of Technology, China), and Guoren Wang (Beijing Institute of Technology, China)

DifFlow3D: Toward Robust Uncertainty-Aware Scene Flow Estimation with Iterative Diffusion-Based Refinement 15109
Jiuming Liu (Shanghai Jiao Tong University, China), Guangming Wang (University of Cambridge, United Kingdom), Weicai Ye (Zhejiang University, China), Chaokang Jiang (PhiGent Robotics, China), Jinru Han (Shanghai Jiao Tong University, China), Zhe Liu (Shanghai Jiao Tong University, China), Guofeng Zhang (Zhejiang University, China), Dalong Du (PhiGent Robotics, China), and Hesheng Wang (Shanghai Jiao Tong University, China)

LMDrive: Closed-Loop End-to-End Driving with Large Language Models 15120
Hao Shao (The Chinese University of Hong Kong, Hong Kong; Sensetime Research, China), Yuxuan Hu (CPII under InnoHK, Hong Kong), Letian Wang (University of Toronto, Canada), Guanglu Song (Sensetime Research, China), Steven L. Waslander (University of Toronto, Canada), Yu Liu (SenseTime Research, China; Shanghai Artificial Intelligence Laboratory, China), and Hongsheng Li (The Chinese University of Hong Kong, Hong Kong; CPII under InnoHK, Hong Kong; Shanghai Artificial Intelligence Laboratory, China)

SOAC: Spatio-Temporal Overlap-Aware Multi-Sensor Calibration using Neural Radiance Fields 15131
Quentin Herau (Huawei Technologies, France), Nathan Piasco (Huawei Technologies, France), Moussab Bennehar (Huawei Technologies, France), Luis Roldao (Huawei Technologies, France), Dzmitry Tsishkou (Huawei Technologies, France), Cyrille Migniot (Université de Bourgogne, France), Pascal Vasseur (Université de Picardie Jules Verne), and Cédric Demonceaux (Université de Bourgogne)

LaMPilot: An Open Benchmark Dataset for Autonomous Driving with Language Model Programs 15141
Yunsheng Ma (Purdue University, USA), Can Cui (Purdue University, USA), Xu Cao (University of Illinois Urbana-Champaign, USA), Wenqian Ye (University of Virginia, USA), Peiran Liu (Purdue University, USA), Juanwu Lu (Purdue University, USA), Amr Abdelraouf (Toyota Motor North America, USA), Rohit Gupta (Toyota Motor North America, USA), Kyungtae Han (Toyota Motor North America, USA), Aniket Bera (Purdue University, USA), James M. Rehg (University of Illinois Urbana-Champaign, USA), and Ziran Wang (Purdue University, USA)

GLiDR: Topologically Regularized Graph Generative Network for Sparse LiDAR Point Clouds . 15152
Prashant Kumar (Indian Institute of Technology, Delhi, India), Kshitij Madhav Bhat (Indian Institute of Technology, Indore, India), Vedang Bhupesh Shenvi Nadkarni (Birla Institute of Technology and Science, Pilani, India), and Prem Kalra (Indian Institute of Technology, Delhi, India)

Towards Robust 3D Object Detection with LiDAR and 4D Radar Fusion in Various Weather Conditions 15162
Yujeong Chae (KAIST, Republic of Korea), Hyeonseong Kim (KAIST, Republic of Korea), and Kuk-Jin Yoon (KAIST, Republic of Korea)

3DSFLabelling: Boosting 3D Scene Flow Estimation by Pseudo Auto-labelling 15173
Chaokang Jiang (PhiGent Robotics, China), Guangming Wang (University of Cambridge, UK), Jiuming Liu (Shanghai Jiaotong University, China), Hesheng Wang (Shanghai Jiaotong University, China), Zhuang Ma (PhiGent Robotics, China), Zhenqiang Liu (PhiGent Robotics, China), Zhujin Liang (PhiGent Robotics, China), Yi Shan (PhiGent Robotics, China), and Dalong Du (PhiGent Robotics, China)

ADA-Track: End-to-End Multi-Camera 3D Multi-Object Tracking with Alternating Detection and Association 15184
Shuxiao Ding (Mercedes-Benz AG, Germany; University of Bonn, Germany), Lukas Schneider (Mercedes-Benz AG, Germany), Marius Cordts (Mercedes-Benz AG, Germany), and Juergen Gall (University of Bonn, Germany; Lamarr Institute for Machine Learning and Artificial Intelligence, Germany)

PointBeV: A Sparse Approach for BeV Predictions 15195
Loick Chambon (ValeoAI and Sorbonne University, France), Eloi Zablocki (ValeoAI, France), Mickaël Chen (ValeoAI, France), Florent Bartoccioni (ValeoAI, France), Patrick Pérez (Kyutai, France), and Matthieu Cord (ValeoAi and Sorbonne University, France)

Light the Night: A Multi-Condition Diffusion Framework for Unpaired Low-Light Enhancement in Autonomous Driving 15205
*Jinlong Li (Cleveland State University), Baolu Li (Cleveland State University), Zhengzhong Tu (University of Texas at Austin), Xinyu Liu (Cleveland State University), Qing Guo (Centre for Frontier AI Research (CFAR), A*STAR), Felix Juefei-Xu (New York University), Runsheng Xu (University of California, Los Angeles), and Hongkai Yu (Cleveland State University)*

CLIP-BEVFormer: Enhancing Multi-View Image-Based BEV Detector with Ground Truth Flow . 15216
Chenbin Pan (Syracuse University), Burhaneddin Yaman (Bosch Research North America & Bosch Center for Artificial Intelligence (BCAI)), Senem Velipasalar (Syracuse University), and Liu Ren (Bosch Research North America & Bosch Center for Artificial Intelligence (BCAI))

Adapting to Length Shift: FlexiLength Network for Trajectory Prediction 15226
Yi Xu (Northeastern University, USA) and Yun Fu (Northeastern University, USA)

UniPAD: A Universal Pre-training Paradigm for Autonomous Driving 15238
Honghui Yang (Zhejiang University, China), Sha Zhang (University of Science and Technology of China, China), Di Huang (The University of Sydney, Australia), Xiaoyang Wu (HongKong University, China), Haoyi Zhu (University of Science and Technology of China, China), Tong He (Shanghai Artificial Intelligence Laboratory, China), Shixiang Tang (Shanghai Artificial Intelligence Laboratory, China), Hengshuang Zhao (HongKong University, China), Qibo Qiu (Zhejiang Lab, China), Binbin Lin (Zhejiang University, China), Xiaofei He (Zhejiang University, China), and Wanli Ouyang (Shanghai Artificial Intelligence Laboratory)

Higher-order Relational Reasoning for Pedestrian Trajectory Prediction 15251
Sungjune Kim (Korea University, Republic of Korea), Hyung-gun Chi (Hanwha Vision America, USA), Hyerin Lim (Hyundai Motor Company, Republic of Korea), Karthik Ramani (Purdue University, USA), Jinkyu Kim (Korea University, Republic of Korea), and Sangpil Kim (Korea University, Republic of Korea)

HPNet: Dynamic Trajectory Forecasting with Historical Prediction Attention 15261
Xiaolong Tang (Institute of Computing Technology, Chinese Academy of Sciences, China), Meina Kan (Institute of Computing Technology, Chinese Academy of Sciences, China), Shiguang Shan (Institute of Computing Technology, Chinese Academy of Sciences, China), Zhilong Ji (Tomorrow Advancing Life, China), Jinfeng Bai (Tomorrow Advancing Life, China), and Xilin Chen (Institute of Computing Technology, Chinese Academy of Sciences, China)

LiSA: LiDAR Localization with Semantic Awareness 15271
Bochun Yang (Xiamen University, China), Zijun Li (Xiamen University, China), Wen Li (Xiamen University, China), Zhipeng Cai (Intel, USA), Chenglu Wen (Xiamen University, China), Yu Zang (Xiamen University, China), Matthias Muller (Intel, USA), and Cheng Wang (Xiamen University, China)

SmartRefine: A Scenario-Adaptive Refinement Framework for Efficient Motion Prediction 15281
Yang Zhou (SenseTime Research, China), Hao Shao (SenseTime Research, China; The Chinese University of Hong Kong, China), Letian Wang (University of Toronto, Canada), Steven L. Waslander (University of Toronto, Canada), Hongsheng Li (The Chinese University of Hong Kong, China; CPII under InnoHK, China; Shanghai Artificial Intelligence Laboratory, China), and Yu Liu (SenseTime Research, China; Shanghai Artificial Intelligence Laboratory, China)

Pseudo Label Refinery for Unsupervised Domain Adaptation on Cross-dataset 3D Object Detection 15291
Zhanwei Zhang (Zhejiang University, China), Minghao Chen (Hangzhou Dianzi University, China), Shuai Xiao (Alibaba Group, China), Liang Peng (Fabu Inc., China), Hengjia Li (Zhejiang University, China), Binbin Lin (Zhejiang University, China), Ping Li (Hangzhou Dianzi University), Wenxiao Wang (Zhejiang University, China), Boxi Wu (Zhejiang University, China), and Deng Cai (Zhejiang University)

Multi-agent Collaborative Perception via Motion-aware Robust Communication Network 15301
Shixin Hong (Shenzhen International Graduate School, Tsinghua University), Yu Liu (Tsinghua University), Zhi Li (Shenzhen International Graduate School, Tsinghua University), Shaohui Li (Shenzhen International Graduate School, Tsinghua University), and You He (Tsinghua University)

TASeg: Temporal Aggregation Network for LiDAR Semantic Segmentation 15311
Xiaopei Wu (State Key Lab of CAD&CG, Zhejiang University, China), Yuenan Hou (Shanghai AI Laboratory, China), Xiaoshui Huang (Shanghai AI Laboratory, China), Binbin Lin (School of Software Technology, Zhejiang University, China), Tong He (Shanghai AI Laboratory, China), Xinge Zhu (The Chinese University of Hong Kong, China), Yuexin Ma (ShanghaiTech University, China), Boxi Wu (School of Software Technology, Zhejiang University, China), Haifeng Liu (State Key Lab of CAD&CG, Zhejiang University, China), Deng Cai (State Key Lab of CAD&CG, Zhejiang University, China), and Wanli Ouyang (Shanghai AI Laboratory, China)

HINTED: Hard Instance Enhanced Detector with Mixed-Density Feature Fusion for Sparsely-Supervised 3D Object Detection 15321
Qiming Xia (Xiamen University, China), Wei Ye (Xiamen University, China), Hai Wu (Xiamen University, China), Shijia Zhao (Xiamen University, China), Leyuan Xing (Xiamen University, China), Xun Huang (Xiamen University, China), Jinhao Deng (Xiamen University, China), Xin Li (Texas A&M University, USA), Chenglu Wen (Xiamen University, China), and Cheng Wang (Xiamen University, China)

CaKDP: Category-aware Knowledge Distillation and Pruning Framework for Lightweight 3D Object Detection 15331
Haonan Zhang (Xi'an Jiaotong University), Longjun Liu (Xi'an Jiaotong University), Yuqi Huang (Xi'an Jiaotong University), Zhao Yang (Xi'an Jiaotong University), Xinyu Lei (Xi'an Jiaotong University), and Bihan Wen (Nanyang Technological University)

Diffusion-ES: Gradient-free Planning with Diffusion for Autonomous and Instruction-guided Driving 15342
Brian Yang (Carnegie Mellon University), Huangyuan Su (Carnegie Mellon University), Nikolaos Gkanatsios (Carnegie Mellon University), Tsung-Wei Ke (Carnegie Mellon University), Ayush Jain (Carnegie Mellon University), Jeff Schneider (Carnegie Mellon University), and Katerina Fragkiadaki (Carnegie Mellon University)

TULIP: Transformer for Upsampling of LiDAR Point Clouds 15354
Bin Yang (ETH Zurich, Switzerland), Patrick Pfreundschuh (ETH Zurich, Switzerland), Roland Siegwart (ETH Zurich, Switzerland), Marco Hutter (ETH Zurich, Switzerland), Peyman Moghadam (CSIRO, Australia), and Vaishakh Patil (ETH Zurich, Switzerland)

Bézier Everywhere All at Once: Learning Drivable Lanes as Bézier Graphs 15365
Hugh Blayney (dRISK.ai, United Kingdom), Hanlin Tian (Imperial College London, United Kingdom), Hamish Scott (dRISK.ai, United Kingdom), Nils Goldbeck (dRISK.ai, United Kingdom), Chess Stetson (dRISK.ai, United Kingdom), and Panagiotis Angeloudis (Imperial College London, United Kingdom)

Flow-Guided Online Stereo Rectification for Wide Baseline Stereo 15375
Anush Kumar (Torc Robotics, USA), Fahim Mannan (Torc Robotics, USA), Omid Hosseini Jafari (Torc Robotics, USA), Shile Li (Torc Robotics, USA), and Felix Heide (Torc Robotics and Princeton University, USA)

LASIL: Learner-Aware Supervised Imitation Learning For Long-term Microscopic Traffic Simulation 15386
Ke Guo (The University of Hong Kong), Zhenwei Miao (Alibaba Group), Wei Jing (Alibaba Group), Weiwei Liu (Alibaba Group), Weizi Li (University of Tennessee, Knoxville), Dayang Hao (Alibaba Group), and Jia Pan (The University of Hong Kong)

HIMap: HybrId Representation Learning for End-to-end Vectorized HD Map Construction 15396
Yi Zhou (Samsung R&D Institute China-Beijing (SRC-B)), Hui Zhang (Samsung R&D Institute China-Beijing (SRC-B)), Jiaqian Yu (Samsung R&D Institute China-Beijing (SRC-B)), Yifan Yang (Samsung R&D Institute China-Beijing (SRC-B)), Sangil Jung (Samsung Advanced Institute of Technology (SAIT), South Korea), Seung-In Park (Samsung Advanced Institute of Technology (SAIT), South Korea), and ByungIn Yoo (Samsung Advanced Institute of Technology (SAIT), South Korea)

RadSimReal: Bridging the Gap Between Synthetic and Real Data in Radar Object Detection With Simulation 15407
Oded Bialer (General Motors) and Yuval Haitman (General Motors; Ben-Gurion University of the Negev, Israel)

3D LiDAR Mapping in Dynamic Environments using a 4D Implicit Neural Representation 15417
Xingguang Zhong (University of Bonn, Germany), Yue Pan (University of Bonn, Germany), Cyrill Stachniss (University of Bonn, Germany), and Jens Behley (University of Bonn, Germany)

Quantifying Uncertainty in Motion Prediction with Variational Bayesian Mixture 15428
Juanwu Lu (Purdue University), Can Cui (Purdue University), Yunsheng Ma (Purdue University), Aniket Bera (Purdue University), and Ziran Wang (Purdue University)

Continual Learning for Motion Prediction Model via Meta-Representation Learning and Optimal Memory Buffer Retention Strategy 15438
DaeJun Kang (Korea Automotive Technology Institute (KATECH), Korea), Dongsuk Kum (Korea Advanced Institute of Science and Technology (KAIST), Korea), and Sanmin Kim (Korea Advanced Institute of Science and Technology (KAIST), Korea)

PARA-Drive: Parallelized Architecture for Real-time Autonomous Driving 15449
Xinshuo Weng (NVIDIA Research), Boris Ivanovic (NVIDIA Research), Yan Wang (NVIDIA Research), Yue Wang (NVIDIA Research, University of Southern California), and Marco Pavone (NVIDIA Research, Stanford University)

ChatScene: Knowledge-Enabled Safety-Critical Scenario Generation for Autonomous Vehicles .. 15459
Jiawei Zhang (University of Illinois Urbana-Champaign), Chejian Xu (University of Illinois Urbana-Champaign), and Bo Li (University of Illinois Urbana-Champaign & University of Chicago)

CRKD: Enhanced Camera-Radar Object Detection with Cross-modality Knowledge Distillation . 15470
Lingjun Zhao (University of Michigan, USA), Jingyu Song (University of Michigan, USA), and Katherine A. Skinner (University of Michigan, USA)

Communication-Efficient Collaborative Perception via Information Filling with Codebook 15481
Yue Hu (Shanghai Jiao Tong University, China), Juntong Peng (Shanghai Jiao Tong University), Sifei Liu (Shanghai Jiao Tong University), Junhao Ge (Shanghai Jiao Tong University), Si Liu (Beihang University), and Siheng Chen (Shanghai Jiao Tong University)

RadarDistill: Boosting Radar-based Object Detection Performance via Knowledge Distillation from LiDAR Features 15491
Geonho Bang (Hanyang University, Korea), Kwangjin Choi (Hanyang University, Korea), Jisong Kim (Hanyang University, Korea), Dongsuk Kum (KAIST, Korea), and Jun Won Choi (Seoul National University, Korea)

ICP-Flow: LiDAR Scene Flow Estimation with ICP 15501
Yancong Lin (Delft University of Technology) and Holger Caesar (Delft University of Technology)

Improving Bird's Eye View Semantic Segmentation by Task Decomposition 15512
Tianhao Zhao (Wuhan University, China), Yongcan Chen (Wuhan University, China), Yu Wu (Wuhan University, China), Tianyang Liu (Wuhan University, China), Bo Du (Wuhan University, China), Peilun Xiao (Didi Chuxing, China), Shi Qiu (Didi Chuxing, China), Hongda Yang (Didi Chuxing, China), Guozhen Li (Didi Chuxing, China), Yi Yang (Didi Chuxing, China), and Yutian Lin (Wuhan University, China)

DriveWorld: 4D Pre-trained Scene Understanding via World Models for Autonomous Driving .. 15522
Chen Min (Peking University), Dawei Zhao (Unmanned Systems Technology Research Center, Defense Innovation Institute), Liang Xiao (Unmanned Systems Technology Research Center, Defense Innovation Institute), Jian Zhao (China Telecom Institute of AI & Northwestern Polytechnical University), Xinli Xu (Hong Kong University of Science and Technology (Guangzhou)), Zheng Zhu (GigaAI), Lei Jin (Beijing University of Posts and Telecommunications), Jianshu Li (Ant Group), Yulan Guo (Sun Yat-sen University), Junliang Xing (Tsinghua University), Liping Jing (Beijing Jiaotong University), Yiming Nie (Unmanned Systems Technology Research Center, Defense Innovation Institute), and Bin Dai (Unmanned Systems Technology Research Center, Defense Innovation Institute)

HRVDA: High-Resolution Visual Document Assistant 15534
Chaohu Liu (University of Science and Technology of China, China), Kun Yin (Tencent YouTu Lab, China), Haoyu Cao (Tencent YouTu Lab, China), Xinghua Jiang (Tencent YouTu Lab, China), Xin Li (Tencent YouTu Lab, China), Yinsong Liu (Tencent YouTu Lab, China), Deqiang Jiang (Tencent YouTu Lab, China), Xing Sun (Tencent YouTu Lab, China), and Linli Xu (University of Science and Technology of China, China)

Enhancing Visual Document Understanding with Contrastive Learning in Large Visual-Language Models 15546
Xin Li (Tencent), Yunfei Wu (Tencent), Xinghua Jiang (Tencent), Zhihao Guo (Tencent), Mingming Gong (Tencent), Haoyu Cao (Tencent), Yinsong Liu (Tencent), Deqiang Jiang (Tencent), and Xing Sun (Tencent)

RoDLA: Benchmarking the Robustness of Document Layout Analysis Models 15556
Yufan Chen (Karlsruhe Institute of Technology, Germany), Jiaming Zhang (Karlsruhe Institute of Technology, Germany), Kunyu Peng (Karlsruhe Institute of Technology, Germany), Junwei Zheng (Karlsruhe Institute of Technology, Germany), Ruiping Liu (Karlsruhe Institute of Technology, Germany), Philip Torr (University of Oxford, UK), and Rainer Stiefelhagen (Karlsruhe Institute of Technology, Germany)

Multi-modal In-Context Learning Makes an Ego-evolving Scene Text Recognizer 15567
Zhen Zhao (East China Normal University, China), Jingqun Tang (ByteDance, China), Chunhui Lin (ByteDance, China), Binghong Wu (ByteDance, China), Can Huang (ByteDance, China), Hao Liu (ByteDance, China), Xin Tan (East China Normal University, China), Zhizhong Zhang (East China Normal University, China), and Yuan Xie (East China Normal University, China)

CMA: A Chromaticity Map Adapter for Robust Detection of Screen-Recapture Document Images 15577
Changsheng Chen (Shenzhen University, China), Liangwei Lin (Shenzhen University, China), Yongqi Chen (Shenzhen University), Bin Li (Shenzhen University), Jishen Zeng (Alibaba Group), and Jiwu Huang (Shenzhen MSU-BIT University)

ODM: A Text-Image Further Alignment Pre-training Approach for Scene Text Detection and Spotting 15587
Chen Duan (Meituan), Pei Fu (Meituan), Shan Guo (Meituan), Qianyi Jiang (Meituan), and Xiaoming Wei (Meituan)

GRAM: Global Reasoning for Multi-Page VQA 15598
Tsachi Blau (Technion Institute of Technology), Sharon Fogel (AWS AI Labs), Roi Ronen (Technion Institute of Technology), Alona Golts (AWS AI Labs), Roy Ganz (Technion Institute of Technology), Elad Ben Avraham (AWS AI Labs), Aviad Aberdam (AWS AI Labs), Shahar Tsiper (AWS AI Labs), and Ron Litman (AWS AI Labs)

Bridging the Gap Between End-to-End and Two-Step Text Spotting 15608
Mingxin Huang (South China University of Technology, China), Hongliang Li (South China University of Technology, China), Yuliang Liu (Huazhong University of Science and Technology, China), Xiang Bai (Huazhong University of Science and Technology, China), and Lianwen Jin (South China University of Technology, China)

An Empirical Study of Scaling Law for Scene Text Recognition 15619
Miao Rang (Huawei Noah's Ark Lab, China), Zhenni Bi (Huawei Noah's Ark Lab, China), Chuanjian Liu (Huawei Noah's Ark Lab, China), Yunhe Wang (Huawei Noah's Ark Lab, China), and Kai Han (Huawei Noah's Ark Lab, China)

LayoutLLM: Layout Instruction Tuning with Large Language Models for Document Understanding... 15630
Chuwei Luo (Alibaba Group), Yufan Shen (Zhejiang University), Zhaoqing Zhu (Alibaba Group), Qi Zheng (Alibaba Group), Zhi Yu (Zhejiang University), and Cong Yao (Alibaba Group)

OmniParser: A Unified Framework for Text Spotting, Key Information Extraction and Table Recognition 15641
Jianqiang Wan (Alibaba Group), Sibo Song (Alibaba Group), Wenwen Yu (Huazhong University of Science and Technology), Yuliang Liu (Huazhong University of Science and Technology), Wenqing Cheng (Huazhong University of Science and Technology), Fei Huang (Alibaba Group), Xiang Bai (Huazhong University of Science and Technology), Cong Yao (Alibaba Group), and Zhibo Yang (Alibaba Group)

DocRes: A Generalist Model Toward Unifying Document Image Restoration Tasks 15654
Jiaxin Zhang (South China University of Technology), Dezhi Peng (South China University of Technology), Chongyu Liu (South China University of Technology), Peirong Zhang (South China University of Technology), and Lianwen Jin (South China University of Technology)

LayoutFormer: Hierarchical Text Detection Towards Scene Text Understanding 15665
Min Liang (University of Science and Technology Beijing, China), Jia-Wei Ma (University of Science and Technology Beijing, China), Xiaobin Zhu (University of Science and Technology Beijing, China), Jingyan Qin (University of Science and Technology Beijing, China), and Xu-Cheng Yin (University of Science and Technology Beijing, China)

Generating Handwritten Mathematical Expressions From Symbol Graphs: An End-to-End Pipeline..... 15675
Yu Chen (Beijing Waiyan Online Digital Technology), Fei Gao (Xidian University, China), Yanguang Zhang (Hangzhou Dianzi University), Maoying Qiao (University of Technology, Sydney (UTS), Australia), and Nannan Wang (Xidian University, China)

OpenESS: Event-based Semantic Scene Understanding with Open Vocabularies 15686
*Lingdong Kong (National University of Singapore), Youquan Liu (Hochschule Bremerhaven), Lai Xing Ng (Institute for Infocomm Research, A*STAR), Benoit R. Cottereau (IPAL, CNRS IRL, Singapore), and Wei Tsang Ooi (National University of Singapore)*

PELA: Learning Parameter-Efficient Models with Low-Rank Approximation 15699
Yangyang Guo (National University of Singapore, Singapore), Guangzhi Wang (National University of Singapore, Singapore), and Mohan Kankanhalli (National University of Singapore)

MADTP: Multimodal Alignment-Guided Dynamic Token Pruning for Accelerating Vision-Language Transformer 15710
Jianjian Cao (Fudan University, China), Peng Ye (Fudan University, China), Shengze Li (Fudan University, China), Chong Yu (Fudan University, China), Yansong Tang (Tsinghua University), Jiwen Lu (Tsinghua University), and Tao Chen (Fudan University)

VkD: Improving Knowledge Distillation using Orthogonal Projections 15720
Roy Miles (Huawei Noah's Ark Lab, UK), Ismail Elezi (Huawei Noah's Ark Lab, UK), and Jiankang Deng (Huawei Noah's Ark Lab, UK)

Logit Standardization in Knowledge Distillation 15731
Shangquan Sun (University of Chinese Academy of Sciences, China; Chinese Academy of Sciences, China), Wenqi Ren (Shenzhen Campus of Sun Yat-sen University, China), Jingzhi Li (Chinese Academy of Sciences, China), Rui Wang (University of Chinese Academy of Sciences, China; Chinese Academy of Sciences, China), and Xiaochun Cao (Shenzhen Campus of Sun Yat-sen University, China)

Multi-criteria Token Fusion with One-step-ahead Attention for Efficient Vision Transformers 15741
Sanghyeok Lee (Korea University, South Korea), Joonmyung Choi (Korea University, South Korea), and Hyunwoo J. Kim (Korea University, South Korea)

ParameterNet: Parameters Are All You Need for Large-scale Visual Pretraining of Mobile Networks 15751
Kai Han (Huawei Noah's Ark Lab), Yunhe Wang (Huawei Noah's Ark Lab), Jianyuan Guo (The University of Sydney), and Enhua Wu (State Key Lab of Computer Science, ISCAS)

DeepCache: Accelerating Diffusion Models for Free 15762
Xinyin Ma (National University of Singapore, Singapore), Gongfan Fang (National University of Singapore, Singapore), and Xinchao Wang (National University of Singapore, Singapore)

ALGM: Adaptive Local-then-Global Token Merging for Efficient Semantic Segmentation with Plain Vision Transformers 15773
Narges Norouzi (Eindhoven University of Technology, Netherlands), Svetlana Orlova (Eindhoven University of Technology, Netherlands), Daan de Geus (Eindhoven University of Technology, Netherlands), and Gijs Dubbelman (Eindhoven University of Technology, Netherlands)

A General and Efficient Training for Transformer via Token Expansion 15783
Wenxuan Huang (East China Normal University, China), Yunhang Shen (Tencent Youtu Lab, China), Jiao Xie (Xiamen University, China), Baochang Zhang (Beihang University, China), Gaoqi He (East China Normal University, China), Ke Li (Tencent Youtu Lab, China), Xing Sun (Tencent Youtu Lab, China), and Shaohui Lin (East China Normal University, China)

Efficient Dataset Distillation via Minimax Diffusion 15793
Jianyang Gu (Zhejiang University, China), Saeed Vahidian (Duke University, USA), Vyacheslav Kungurtsev (Czech Technical University, Czech Republic), Haonan Wang (National University of Singapore, Singapore), Wei Jiang (Zhejiang University, China), Yang You (National University of Singapore, Singapore), and Yiran Chen (Duke University, USA)

PEM: Prototype-based Efficient MaskFormer for Image Segmentation 15804
Niccolò Cavagnero (Politecnico di Torino, Italy), Gabriele Rosi (Politecnico di Torino, Italy), Claudia Cuttano (Politecnico di Torino, Italy), Francesca Pistilli (Politecnico di Torino, Italy), Marco Ciccone (Politecnico di Torino, Italy), Giuseppe Averta (Politecnico di Torino, Italy), and Fabio Cermelli (Focoos AI, Italy)

Transferable and Principled Efficiency for Open-Vocabulary Segmentation 15814
Jingxuan Xu (Beijing Jiaotong University), Wuyang Chen (Simon Fraser University), Yao Zhao (Beijing Jiaotong University), and Yunchao Wei (Beijing Jiaotong University)

Dense Vision Transformer Compression with Few Samples 15825
Hanxiao Zhang (Nanjing University, China), Yifan Zhou (Nanjing University, China), and Guo-Hua Wang (Nanjing University, China)

Dr2Net: Dynamic Reversible Dual-Residual Networks for Memory-Efficient Finetuning 15835
Chen Zhao (King Abdullah University of Science and Technology, Saudi Arabia), Shuming Liu (King Abdullah University of Science and Technology, Saudi Arabia), Karttikeya Mangalam (UC Berkeley, USA), Guocheng Qian (King Abdullah University of Science and Technology, Saudi Arabia), Fatimah Zohra (King Abdullah University of Science and Technology, Saudi Arabia), Abdulmohsen Alghannam (King Abdullah University of Science and Technology, Saudi Arabia), Jitendra Malik (UC Berkeley, USA), and Bernard Ghanem (King Abdullah University of Science and Technology, Saudi Arabia)

MaxQ: Multi-Axis Query for N:M Sparsity Network 15845
Jingyang Xiang (Zhejiang University, China), Siqi Li (Zhejiang University, China), Junhao Chen (Zhejiang University, China), Zhuangzhi Chen (Zhejiang University of Technology, China), Tianxin Huang (Zhejiang University, China), Linpeng Peng (Zhejiang University, China), and Yong Liu (Zhejiang University, China)

Retraining-Free Model Quantization via One-Shot Weight-Coupling Learning 15855
Chen Tang (Tsinghua University, China), Yuan Meng (Tsinghua University, China), Jiacheng Jiang (Tsinghua University, China), Shuzhao Xie (Tsinghua University, China), Rongwei Lu (Tsinghua University, China), Xinzhu Ma (The Chinese University of Hong Kong, Hong Kong), Zhi Wang (Tsinghua University, China), and Wenwu Zhu (Tsinghua University, China)

LORS: Low-rank Residual Structure for Parameter-Efficient Network Stacking 15866
Jialin Li (Tencent YouTu Lab, China), Qiang Nie (Tencent YouTu Lab, China), Weifu Fu (Tencent YouTu Lab, China), Yuhuan Lin (Tencent YouTu Lab, China), Guangpin Tao (Tencent YouTu Lab, China), Yong Liu (Tencent YouTu Lab, China), and Chengjie Wang (Tencent YouTu Lab, China)

Towards High-fidelity Artistic Image Vectorization via Texture-Encapsulated Shape Parameterization 15877
Ye Chen (Shanghai Jiao Tong University, China), Bingbing Ni (Shanghai Jiao Tong University, China; USC-SJTU Institute of Cultural and Creative Industry, China), Jinfan Liu (Shanghai Jiao Tong University, China), Xiaoyang Huang (Shanghai Jiao Tong University, China), and Xuanhong Chen (Shanghai Jiao Tong University, China; USC-SJTU Institute of Cultural and Creative Industry, China)

Learning Vision from Models Rivals Learning Vision from Data 15887
Yonglong Tian (Google Research), Lijie Fan (MIT CSAIL), Kaifeng Chen (Google Research), Dina Katabi (MIT CSAIL), Dilip Krishnan (Google Research), and Phillip Isola (MIT CSAIL)

Efficient Multitask Dense Predictor via Binarization 15899
Yuzhang Shang (Illinois Institute of Technology), Dan Xu (HKUST), Gaowen Liu (CISCO), Ramana Rao Kompella (CISCO), and Yan Yan (Illinois Institute of Technology)

RepViT: Revisiting Mobile CNN From ViT Perspective 15909
Ao Wang (Tsinghua University, BNRist, China), Hui Chen (Tsinghua University, BNRist, China), Zijia Lin (Tsinghua University, China), Jungong Han (The University of Sheffield), and Guiguang Ding (Tsinghua University, BNRist, China)

Enhancing Post-training Quantization Calibration through Contrastive Learning 15921
Yuzhang Shang (Illinois Institute of Technology), Gaowen Liu (CISCO), Ramana Rao Kompella (CISCO), and Yan Yan (Illinois Institute of Technology)

FreeKD: Knowledge Distillation via Semantic Frequency Prompt 15931
Yuan Zhang (Peking University), Tao Huang (The University of Sydney), Jiaming Liu (Peking University), Tao Jiang (Zhejiang University), Kuan Cheng (Peking University), and Shanghang Zhang (Peking University)

PTQ4SAM: Post-Training Quantization for Segment Anything 15941
Chengtao Lv (Beihang University, China), Hong Chen (Beihang University, China), Jinyang Guo (Beihang University, China), Yifu Ding (Beihang University, China), and Xianglong Liu (Beihang University, China; Zhongguancun Laboratory, China; Hefei Comprehensive National Science Center, China)

CLIP-KD: An Empirical Study of CLIP Model Distillation 15952
Chuanguang Yang (Institute of Computing Technology, Chinese Academy of Sciences), Zhulin An (Institute of Computing Technology, Chinese Academy of Sciences), Libo Huang (Institute of Computing Technology, Chinese Academy of Sciences), Junyu Bi (Institute of Computing Technology, Chinese Academy of Sciences), Xinqiang Yu (Institute of Computing Technology, Chinese Academy of Sciences), Han Yang (Institute of Computing Technology, Chinese Academy of Sciences), Boyu Diao (Institute of Computing Technology, Chinese Academy of Sciences), and Yongjun Xu (Institute of Computing Technology, Chinese Academy of Sciences)

MobileCLIP: Fast Image-Text Models through Multi-Modal Reinforced Training 15963
Pavan Kumar Anasosalu Vasu (Apple, USA), Hadi Pouransari (Apple, USA), Fartash Faghri (Apple, USA), Raviteja Vemulapalli (Apple, USA), and Oncel Tuzel (Apple, USA)

Scaled Decoupled Distillation 15975
Shicai Wei (University of Electronic Science and Technology of China), Chunbo Luo (University of Electronic Science and Technology of China), and Yang Luo (University of Electronic Science and Technology of China)

Self-Distilled Masked Auto-Encoders are Efficient Video Anomaly Detectors 15984
Nicolae-Cătălin Ristea (University of Bucharest), Florinel-Alin Croitoru (University of Bucharest), Radu Tudor Ionescu (University of Bucharest), Marius Popescu (University of Bucharest), Fahad Shahbaz Khan (MBZ University of Artificial Intelligence), and Mubarak Shah (University of Central Florida)

PikeLPN: Mitigating Overlooked Inefficiencies of Low-Precision Neural Networks 15996
Marina Neseem (Brown University, USA), Conor McCullough (Google, USA), Randy Hsin (Google, USA), Chas Leichner (Google, USA), Shan Li (Google, USA), In Suk Chong (Google, USA), Andrew Howard (Google, USA), Lukasz Lew (Google, USA), Sherief Reda (Brown University, USA), Ville-Mikko Rautio (Google, USA), and Daniele Moro (Google, USA)

C2KD: Bridging the Modality Gap for Cross-Modal Knowledge Distillation 16006
Fushuo Huo (The Hong Kong Polytechnic University, Hong Kong SAR, China), Wenchao Xu (The Hong Kong Polytechnic University, Hong Kong SAR, China), Jingcai Guo (The Hong Kong Polytechnic University, Hong Kong SAR, China), Haozhao Wang (Huazhong University of Science and Technology, China), and Song Guo (The Hong Kong University of Science and Technology, Hong Kong SAR, China)

KD-DETR: Knowledge Distillation for Detection Transformer with Consistent Distillation Points Sampling 16016
Yu Wang (Baidu VIS, China), Xin Li (Baidu VIS, China), Shengzhao Weng (Baidu VIS, China), Gang Zhang (Baidu VIS, China), Haixiao Yue (Baidu VIS, China), Haocheng Feng (Baidu VIS, China), Junyu Han (Baidu VIS, China), and Errui Ding (Baidu VIS, China)

Towards Accurate Post-training Quantization for Diffusion Models 16026
Changyuan Wang (Tsinghua University, China), Ziwei Wang (Carnegie Mellon University, USA), Xiuwei Xu (Tsinghua University, China), Yansong Tang (Tsinghua University, China), Jie Zhou (Tsinghua University, China), and Jiwen Lu (Tsinghua University, China)

CURSOR: Scalable Mixed-Order Hypergraph Matching with CUR Decomposition 16036
Qixuan Zheng (City University of Hong Kong, Hong Kong), Ming Zhang (Hong Kong Applied Science and Technology Research Institute, Hong Kong), and Hong Yan (City University of Hong Kong, Hong Kong)

Frozen Feature Augmentation for Few-Shot Image Classification 16046
Andreas Bär (Technische Universität Braunschweig, Germany), Neil Houlsby (Google DeepMind, Switzerland), Mostafa Dehghani (Google DeepMind, USA), and Manoj Kumar (Google DeepMind, Netherlands)

Jointly Training and Pruning CNNs via Learnable Agent Guidance and Alignment 16058
Alireza Ganjdanesh (University of Maryland College Park), Shangqian Gao (University of Pittsburgh), and Heng Huang (University of Maryland College Park)

Zero-TPrune: Zero-Shot Token Pruning through Leveraging of the Attention Graph in Pre-Trained Transformers 16070
Hongjie Wang (Princeton University, USA), Bhishma Dedhia (Princeton University, USA), and Niraj K. Jha (Princeton University, USA)

Attention-Driven Training-Free Efficiency Enhancement of Diffusion Models 16080
Hongjie Wang (Princeton University, USA), Difan Liu (Adobe Research, USA), Yan Kang (Adobe Research, USA), Yijun Li (Adobe Research, USA), Zhe Lin (Adobe Research, USA), Niraj K. Jha (Princeton University, USA), and Yuchen Liu (Adobe Research, USA)

BilevelPruning: Unified Dynamic and Static Channel Pruning for Convolutional Neural Networks 16090
Shangqian Gao (University of Pittsburgh, USA), Yanfu Zhang (College of William and Mary, USA), Feihu Huang (University of Pittsburgh, USA), and Heng Huang (University of Maryland College Park, USA)

Low-Rank Rescaled Vision Transformer Fine-Tuning: A Residual Design Approach 16101
Wei Dong (University of Electronic Science and Technology of China), Xing Zhang (Xi'an University of Architecture and Technology), Bihui Chen (Xi'an University of Architecture and Technology), Dawei Yan (Xi'an University of Architecture and Technology), Zhijun Lin (Northwestern Polytechnical University), Qingsen Yan (Northwestern Polytechnical University), Peng Wang (University of Electronic Science and Technology of China), and Yang Yang (University of Electronic Science and Technology of China)

EfficientSAM: Leveraged Masked Image Pretraining for Efficient Segment Anything 16111
Yunyang Xiong (Meta AI Research), Bala Varadarajan (Meta AI Research), Lemeng Wu (Meta AI Research), Xiaoyu Xiang (Meta AI Research), Fanyi Xiao (Meta AI Research), Chenchen Zhu (Meta AI Research), Xiaoliang Dai (Meta AI Research), Dilin Wang (Meta AI Research), Fei Sun (Meta AI Research), Forrest Iandola (Meta AI Research), Raghuraman Krishnamoorthi (Meta AI Research), and Vikas Chandra (Meta AI Research)

FlashEval: Towards Fast and Accurate Evaluation of Text-to-image Diffusion Generative Models 16122
Lin Zhao (Infinigence-AI), Tianchen Zhao (Tsinghua University), Zinan Lin (Microsoft Research), Xuefei Ning (Tsinghua University), Guohao Dai (Shanghai Jiao Tong University), Huazhong Yang (Tsinghua University), and Yu Wang (Tsinghua University)

Instance-Aware Group Quantization for Vision Transformers 16132
Jaehyeon Moon (Yonsei University, South Korea; Articron, South Korea), Dohyung Kim (Yonsei University, South Korea), Junyong Cheon (Yonsei University, South Korea), and Bumsub Ham (Yonsei University, South Korea)

Finding Lottery Tickets in Vision Models via Data-driven Spectral Foresight Pruning 16142
Leonardo Iurada (Politecnico di Torino, Italy), Marco Ciccone (Politecnico di Torino, Italy), and Tatiana Tommasi (Politecnico di Torino, Italy)

Joint-Task Regularization for Partially Labeled Multi-Task Learning 16152
Kento Nishi (Harvard University), Junsik Kim (Harvard University), Wanhua Li (Harvard University), and Hanspeter Pfister (Harvard University)

Auto-Train-Once: Controller Network Guided Automatic Network Pruning from Scratch 16163
Xidong Wu (University of Pittsburgh, USA), Shangqian Gao (University of Pittsburgh, USA), Zeyu Zhang (University of Arizona, USA), Zhenzhen Li (Bosch Center for AI, USA), Runxue Bao (GE HealthCare, USA), Yanfu Zhang (College of William and Mary, USA), Xiaoqian Wang (Purdue University, USA), and Heng Huang (University of Maryland College Park, USA)

Reg-PTQ: Regression-specialized Post-training Quantization for Fully Quantized Object Detector 16174
Yifu Ding (Beihang University, China), Weilun Feng (Beihang University, China), Chuyan Chen (Beihang University, China), Jinyang Guo (Beihang University, China), and Xianglong Liu (Beihang University, China; Zhongguancun Laboratory, China; Hefei Comprehensive National Science Center, China)

MULTIFLOW: Shifting Towards Task-Agnostic Vision-Language Pruning 16185
Matteo Farina (University of Trento, Italy), Massimiliano Mancini (University of Trento, Italy), Elia Cunegatti (University of Trento, Italy), Gaowen Liu (Cisco Research, USA), Giovanni Iacca (University of Trento, Italy), and Elisa Ricci (University of Trento, Italy)

MTLoRA: Low-Rank Adaptation Approach for Efficient Multi-Task Learning 16196
Ahmed Agiza (Brown University), Marina Neseem (Brown University), and Sherief Reda (Brown University)

Resource-Efficient Transformer Pruning for Finetuning of Large Models 16206
Fatih Ilhan (Georgia Institute of Technology, USA), Gong Su (IBM Research, USA), Selim Furkan Tekin (Georgia Institute of Technology, USA), Tiansheng Huang (Georgia Institute of Technology, USA), Sihao Hu (Georgia Institute of Technology, USA), and Ling Liu (Georgia Institute of Technology, USA)

Promptable Behaviors: Personalizing Multi-Objective Rewards from Human Preferences 16216
Minyoung Hwang (Carnegie Mellon University, Allen Institute for AI), Luca Weihs (Allen Institute for AI), Chanwoo Park (Massachusetts Institute of Technology), Kimin Lee (Korea Advanced Institute of Science and Technology), Aniruddha Kembhavi (Allen Institute for AI), and Kiana Ehsani (Allen Institute for AI)

Holodeck: Language Guided Generation of 3D Embodied AI Environments 16227
Yue Yang (University of Pennsylvania), Fan-Yun Sun (Stanford University), Luca Weihs (Allen Institute for Artificial Intelligence), Eli VanderBilt (Allen Institute for Artificial Intelligence), Alvaro Herrasti (Allen Institute for Artificial Intelligence), Winson Han (Allen Institute for Artificial Intelligence), Jiajun Wu (Stanford University), Nick Haber (Stanford University), Ranjay Krishna (University of Washington), Lingjie Liu (University of Pennsylvania), Chris Callison-Burch (University of Pennsylvania), Mark Yatskar (University of Pennsylvania), Aniruddha Kembhavi (Allen Institute for Artificial Intelligence), and Christopher Clark (Allen Institute for Artificial Intelligence)

SPOC: Imitating Shortest Paths in Simulation Enables Effective Navigation and Manipulation in the Real World 16238
Kiana Ehsani (Allen Institute for AI, USA), Tanmay Gupta (Allen Institute for AI, USA), Rose Hendrix (Allen Institute for AI, USA), Jordi Salvador (Allen Institute for AI, USA), Luca Weihs (Allen Institute for AI, USA), Kuo-Hao Zeng (Allen Institute for AI, USA), Kunal Pratap Singh (EPFL, Switzerland), Yejin Kim (Allen Institute for AI, USA), Winson Han (Allen Institute for AI, USA), Alvaro Herrasti (Allen Institute for AI, USA), Ranjay Krishna (Allen Institute for AI, University of Washington, USA), Dustin Schwenk (Allen Institute for AI, USA), Eli VanderBilt (Allen Institute for AI, USA), and Aniruddha Kembhavi (Allen Institute for AI, USA)

RILA: Reflective and Imaginative Language Agent for Zero-Shot Semantic Audio-Visual Navigation 16251
Zeyuan Yang (Tsinghua University, China), Jiageng Liu (Zhejiang University, China), Peihao Chen (South China University of Technology, China), Anoop Cherian (Mitsubishi Electric Research Laboratories (MERL), America), Tim K. Marks (Mitsubishi Electric Research Laboratories (MERL), America), Jonathan Le Roux (Mitsubishi Electric Research Laboratories (MERL), America), and Chuang Gan (UMass Amherst, USA; MIT-IBM AI Lab, USA)

PhyScene: Physically Interactable 3D Scene Synthesis for Embodied AI 16262
Yandan Yang (National Key Laboratory of General Artificial Intelligence, China), Baoxiong Jia (National Key Laboratory of General Artificial Intelligence, China), Peiyuan Zhi (National Key Laboratory of General Artificial Intelligence, China), and Siyuan Huang (National Key Laboratory of General Artificial Intelligence, China)

Seeing the Unseen: Visual Common Sense for Semantic Placement 16273
Ram Ramrakhya (Georgia Institute of Technology), Aniruddha Kembhavi (Allen Institute of AI), Dhruv Batra (Georgia Institute of Technology), Zsolt Kira (Georgia Institute of Technology), Kuo-Hao Zeng (Allen Institute of AI), and Luca Weihs (Allen Institute of AI)

LEMON: Learning 3D Human-Object Interaction Relation from 2D Images 16284
Yuhang Yang (University of Science and Technology of China), Wei Zhai (University of Science and Technology of China), Hongchen Luo (University of Science and Technology of China), Yang Cao (University of Science and Technology of China), and Zheng-Jun Zha (University of Science and Technology of China)

OVER-NAV: Elevating Iterative Vision-and-Language Navigation with Open-Vocabulary Detection and StructurEd Representation 16296
Ganlong Zhao (The University of Hong Kong; Sun Yat-sen University), Guanbin Li (Sun Yat-sen University; GuangDong Province Key Laboratory of Information Security Technology), Weikai Chen (Digital Content Technology Center, Tencent Games), and Yizhou Yu (The University of Hong Kong)

MP5: A Multi-modal Open-ended Embodied System in Minecraft via Active Perception 16307
Yiran Qin (The Chinese University of Hong Kong, Shenzhen), Enshen Zhou (BeiHang University), Qichang Liu (Tsinghua University), Zhenfei Yin (The University of Sydney), Lu Sheng (Beihang University), Ruimao Zhang (The Chinese University of Hong Kong, Shenzhen), Yu Qiao (Shanghai Artificial Intelligence Laboratory), and Jing Shao (Shanghai Artificial Intelligence Laboratory)

Volumetric Environment Representation for Vision-Language Navigation 16317
Rui Liu (Zhejiang University), Wenguan Wang (Zhejiang University), and Yi Yang (Zhejiang University)

Instance-aware Exploration-Verification-Exploitation for Instance ImageGoal Navigation 16329
Xiaohan Lei (University of Science and Technology of China), Min Wang (Hefei Comprehensive National Science Center), Wengang Zhou (Univerisity of Science and Technology of China), Li Li (Univerisity of Science and Technology of China), and Houqiang Li (Univerisity of Science and Technology of China)

UniGarmentManip: A Unified Framework for Category-Level Garment Manipulation via Dense Visual Correspondence 16340
Ruihai Wu (Peking University, China), Haoran Lu (Peking University, China), Yiyan Wang (Beijing Institute of Technology, China), Yubo Wang (Peking University, China), and Hao Dong (Peking University, China)

Evidential Active Recognition: Intelligent and Prudent Open-World Embodied Perception 16351
Lei Fan (Northwestern University, USA), Mingfu Liang (Northwestern University, USA), Yunxuan Li (Northwestern University, USA), Gang Hua (Wormpex AI Research, USA), and Ying Wu (Northwestern University, USA)

GenH2R: Learning Generalizable Human-to-Robot Handover via Scalable Simulation, Demonstration, and Imitation 16362
Zifan Wang (Tsinghua University, Shanghai Qi Zhi Institute), Junyu Chen (Tsinghua University, Shanghai Qi Zhi Institute), Ziqing Chen (Tsinghua University), Pengwei Xie (Tsinghua University), Rui Chen (Tsinghua University), and Li Yi (Tsinghua University, Shanghai Artificial Intelligence Laboratory, Shanghai Qi Zhi Institute)

GOAT-Bench: A Benchmark for Multi-Modal Lifelong Navigation 16373
Mukul Khanna (Georgia Tech), Ram Ramrakhya (Georgia Tech), Gunjan Chhablani (Georgia Tech), Sriram Yenamandra (Georgia Tech), Theophile Gervet (Carnegie Mellon University), Matthew Chang (University of Illinois Urbana-Champaign), Zsolt Kira (Georgia Tech), Devendra Singh Chaplot (Mistral AI), Dhruv Batra (Georgia Tech), and Roozbeh Mottaghi (University of Washington)

Habitat Synthetic Scenes Dataset (HSSD-200): An Analysis of 3D Scene Scale and Realism Tradeoffs for ObjectGoal Navigation 16384
Mukul Khanna (Georgia Institute of Technology), Yongsen Mao (Simon Fraser University), Hanxiao Jiang (Simon Fraser University), Sanjay Haresh (Simon Fraser University), Brennan Shacklett (Stanford University), Dhruv Batra (Georgia Institute of Technology, Meta AI), Alexander Clegg (Meta AI), Eric Undersander (Meta AI), Angel X. Chang (Simon Fraser University), and Manolis Savva (Simon Fraser University)

Active Open-Vocabulary Recognition: Let Intelligent Moving Mitigate CLIP Limitations 16394
Lei Fan (Northwestern University, USA), Jianxiong Zhou (Northwestern University, USA), Xiaoying Xing (Northwestern University, USA), and Ying Wu (Northwestern University, USA)

Rapid Motor Adaptation for Robotic Manipulator Arms 16404
Yichao Liang (Univerisity of Cambridge, UK), Kevin Ellis (Cornell University, US), and João Henriques (University of Oxford, UK)

Imagine Before Go: Self-Supervised Generative Map for Object Goal Navigation 16414
Sixian Zhang (Institute of Computing Technology, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), Xinyao Yu (Institute of Computing Technology, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), Xinhang Song (Institute of Computing Technology, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), Xiaohan Wang (Institute of Computing Technology, Chinese Academy of Sciences, China), and Shuqiang Jiang (Institute of Computing Technology, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China; Institute of Intelligent Computing Technology, Suzhou, CAS, China)

Auto MC-Reward: Automated Dense Reward Design with Large Language Models for Minecraft 16426
Hao Li (Chinese University of Hong Kong, Hong Kong), Xue Yang (Shanghai AI Laboratory, China), Zhaokai Wang (Shanghai AI Laboratory, China), Xizhou Zhu (Tsinghua University, China), Jie Zhou (Tsinghua University, China), Yu Qiao (Shanghai AI Laboratory, China), Xiaogang Wang (Chinese University of Hong Kong, Hong Kong), Hongsheng Li (Chinese University of Hong Kong, Hong Kong), Lewei Lu (Sensetime Research, China), and Jifeng Dai (Tsinghua University, China)

GenNBV: Generalizable Next-Best-View Policy for Active 3D Reconstruction 16436
Xiao Chen (Shanghai AI Laboratory, China; The Chinese University of Hong Kong, China), Quanyi Li (Shanghai AI Laboratory, China), Tai Wang (Shanghai AI Laboratory), Tianfan Xue (The Chinese University of Hong Kong), and Jiangmiao Pang (Shanghai AI Laboratory)

An Interactive Navigation Method with Effect-oriented Affordance 16446
Xiaohan Wang (Xi'an Jiaotong University, China), Yuehu Liu (Xi'an Jiaotong University, China), Xinhang Song (Institute of Computing Technology, Chinese Academy of Sciences, China), Yuyi Liu (Institute of Computing Technology, University of the Chinese Academy of Sciences, China), Sixian Zhang (Institute of Computing Technology, Chinese Academy of Sciences, China), and Shuqiang Jiang (Institute of Computing Technology, Chinese Academy of Sciences, China)

A Category Agnostic Model for Visual Rearrangment 16457
Yuyi Liu (Institute of Computing Technology, China; University of Chinese Academy of Sciences, China), Xinhang Song (Institute of Computing Technology, China; University of Chinese Academy of Sciences, China), Weijie Li (Institute of Computing Technology, China; University of Chinese Academy of Sciences, China), Xiaohan Wang (Institute of Computing Technology, China), and Shuqiang Jiang (Institute of Computing Technology, China; University of Chinese Academy of Sciences, China)

SkillDiffuser: Interpretable Hierarchical Planning via Skill Abstractions in Diffusion-Based Task Execution 16467
Zhixuan Liang (The University of Hong Kong, Hong Kong SAR), Yao Mu (The University of Hong Kong, Hong Kong SAR), Hengbo Ma (University of California, Berkeley, USA), Masayoshi Tomizuka (University of California, Berkeley, USA), Mingyu Ding (University of California, Berkeley, USA), and Ping Luo (The University of Hong Kong, Hong Kong SAR)

Fusing Personal and Environmental Cues for Identification and Segmentation of First-Person Camera Wearers in Third-Person Views 16477
Ziwei Zhao (Indiana University Bloomington), Yuchen Wang (Indiana University Bloomington), and Chuhua Wang (Indiana University Bloomington)

OpenEQA: Embodied Question Answering in the Era of Foundation Models 16488
Arjun Majumdar (Georgia Institute of Technology), Anurag Ajay (Massachusetts Institute of Technology), Xiaohan Zhang (State University of New York at Binghamton), Pranav Putta (Georgia Institute of Technology), Sriram Yenamandra (Georgia Institute of Technology), Mikael Henaff (Meta), Sneha Silwal (Meta), Paul Mcvay (Meta), Oleksandr Maksymets (Meta), Sergio Arnaud (Meta), Karmesh Yadav (Meta), Qiyang Li (University of California, Berkeley), Ben Newman (Carnegie Mellon Univeristy), Mohit Sharma (Carnegie Mellon Univeristy), Vincent Berges (Meta), Shiqi Zhang (State University of New York at Binghamton), Pulkit Agrawal (Massachusetts Institute of Technology), Yonatan Bisk (Meta), Dhruv Batra (Meta), Mrinal Kalakrishnan (Meta), Franziska Meier (Meta), Chris Paxton (Meta), Alexander Sax (Meta), and Aravind Rajeswaran (Meta)

Model Adaptation for Time Constrained Embodied Control 16499
Jaehyun Song (Sungkyunkwan University, Korea), Minjong Yoo (Sungkyunkwan University, Korea), and Honguk Woo (Sungkyunkwan University, Korea)

You'll Never Walk Alone: A Sketch and Text Duet for Fine-Grained Image Retrieval 16509
Subhadeep Koley (University of Surrey, United Kingdom; iFlyTek-Surrey Joint Research Centre on Artificial Intelligence), Ayan Kumar Bhunia (University of Surrey, United Kingdom), Aneeshan Sain (University of Surrey, United Kingdom), Pinaki Nath Chowdhury (University of Surrey, United Kingdom), Tao Xiang (University of Surrey, United Kingdom; iFlyTek-Surrey Joint Research Centre on Artificial Intelligence), and Yi-Zhe Song (University of Surrey, United Kingdom; iFlyTek-Surrey Joint Research Centre on Artificial Intelligence)

CrossKD: Cross-Head Knowledge Distillation for Object Detection 16520
Jiabao Wang (Nankai University, China), Yuming Chen (Nankai University, China), Zhaohui Zheng (Nankai University, China), Xiang Li (Nankai University, China), Ming-Ming Cheng (Nankai University, China), and Qibin Hou (Nankai University, China)

ProTeCt: Prompt Tuning for Taxonomic Open Set Classification 16531
Tz-Ying Wu (University of California San Diego), Chih-Hui Ho (University of California San Diego), and Nuno Vasconcelos (University of California San Diego)

CAT: Exploiting Inter-Class Dynamics for Domain Adaptive Object Detection 16541
*Mikhail Kennerley (National University of Singapore, Singapore), Jian-Gang Wang (A*STAR, Singapore), Bharadwaj Veeravalli (National University of Singapore, Singapore), and Robby T. Tan (ASUS Intelligent Cloud Services, Singapore)*

Text Is MASS: Modeling as Stochastic Embedding for Text-Video Retrieval 16551
Jiamian Wang (Rochester Institute of Technology), Guohao Sun (Rochester Institute of Technology), Pichao Wang (Amazon Prime Video), Dongfang Liu (Rochester Institute of Technology), Sohail Dianat (Rochester Institute of Technology), Majid Rabbani (Rochester Institute of Technology), Raghuveer Rao (Army Research Laboratory), and Zhiqiang Tao (Rochester Institute of Technology)

UniMODE: Unified Monocular 3D Object Detection 16561
Zhuoling Li (The University of Hong Kong), Xiaogang Xu (The Chinese University of Hong Kong), SerNam Lim (University of Central Florida), and Hengshuang Zhao (The University of Hong Kong)

OVMR: Open-Vocabulary Recognition with Multi-Modal References 16571
Zehong Ma (Peking University), Shiliang Zhang (Peking University), Longhui Wei (Huawei Inc.), and Qi Tian (Huawei Inc.)

From Isolated Islands to Pangea: Unifying Semantic Space for Human Action Understanding 16582
Yong-Lu Li (Shanghai Jiao Tong University, China), Xiaoqian Wu (Shanghai Jiao Tong University, China), Xinpeng Liu (Shanghai Jiao Tong University, China), Zehao Wang (Shanghai Jiao Tong University, China), Yiming Dou (Shanghai Jiao Tong University, China), Yikun Ji (Shanghai Jiao Tong University, China), Junyi Zhang (Shanghai Jiao Tong University, China), Yixing Li (Shanghai Jiao Tong University, China), Xudong Lu (Shanghai Jiao Tong University, China), Jingru Tan (Shanghai Jiao Tong University, China), and Cewu Lu (Shanghai Jiao Tong University, China)

Language-conditioned Detection Transformer 16593
Jang Hyun Cho (UT Austin) and Philipp Krähenbühl (UT Austin)

Distribution-aware Knowledge Prototyping for Non-exemplar Lifelong Person Re-identification 16604
Kunlun Xu (Peking University, China), Xu Zou (Huazhong University of Science and Technology, China), Yuxin Peng (Peking University, China), and Jiahuan Zhou (Peking University, China)

Learning Continual Compatible Representation for Re-indexing Free Lifelong Person Re-identification 16614
Zhenyu Cui (Peking University), Jiahuan Zhou (Peking University), Xun Wang (ByteDance Inc), Manyu Zhu (ByteDance Inc), and Yuxin Peng (Peking University)

Active Object Detection with Knowledge Aggregation and Distillation from Large Models 16624
Dejie Yang (Peking University, China) and Yang Liu (Peking University, China)

SHiNe: Semantic Hierarchy Nexus for Open-vocabulary Object Detection 16634
Mingxuan Liu (University of Trento, NAVER LABS Europe), Tyler L. Hayes (NAVER LABS Europe), Elisa Ricci (University of Trento, Fondazione Bruno Kessler), Gabriela Csurka (NAVER LABS Europe), and Riccardo Volpi (NAVER LABS Europe)

Object Recognition as Next Token Prediction 16645
Kaiyu Yue (University of Maryland, USA), Bor-Chun Chen (Meta, USA), Jonas Geiping (ELLIS Institute Tübingen, Germany), Hengduo Li (Meta, USA), Tom Goldstein (University of Maryland, USA), and Ser-Nam Lim (University of Central Florida)

Exploring the Potential of Large Foundation Models for Open-Vocabulary HOI Detection 16657
Ting Lei (Peking University, China), Shaofeng Yin (Peking University, China), and Yang Liu (Peking University, China)

Gradient Reweighting: Towards Imbalanced Class-Incremental Learning 16668
Jiangpeng He (Purdue University)

Learning Background Prompts to Discover Implicit Knowledge for Open Vocabulary Object Detection 16678
Jiaming Li (Sun Yat-sen University, China), Jiacheng Zhang (Sun Yat-sen University, China), Jichang Li (The University of Hong Kong, Hong Kong), Ge Li (Peking University, China), Si Liu (Beihang University, China), Liang Lin (Sun Yat-sen University), and Guanbin Li (Sun Yat-sen University)

Multi-View Attentive Contextualization for Multi-View 3D Object Detection 16688
Xianpeng Liu (North Carolina State University, USA), Ce Zheng (University of Central Florida, USA), Ming Qian (Ant Group, China), Nan Xue (Ant Group, China), Chen Chen (University of Central Florida, USA), Zhebin Zhang (OPPO USA Research Center, USA), Chen Li (OPPO USA Research Center, USA), and Tianfu Wu (North Carolina State University, USA)

RealNet: A Feature Selection Network with Realistic Synthetic Anomaly for Anomaly Detection 16699
Ximiao Zhang (Capital Normal University, China), Min Xu (Capital Normal University, China), and Xiuzhuang Zhou (Beijing University of Posts and Telecommunications, China)

Generalized Large-Scale Data Condensation via Various Backbone and Statistical Matching 16709
Shitong Shao (Mohamed bin Zayed University of AI, UAE), Zeyuan Yin (Mohamed bin Zayed University of AI, UAE), Muxin Zhou (Mohamed bin Zayed University of AI, UAE), Xindong Zhang (OPPO Research, Guang dong, China), and Zhiqiang Shen (Mohamed bin Zayed University of AI, UAE)

Unleashing Unlabeled Data: A Paradigm for Cross-View Geo-Localization 16719
Guopeng Li (Wuhan University), Ming Qian (State Key Lab. LIESMARS, Wuhan University), and Gui-Song Xia (State Key Lab. LIESMARS, Wuhan University)

PointOBB: Learning Oriented Object Detection via Single Point Supervision 16730
Junwei Luo (Wuhan University), Xue Yang (Shanghai AI Laboratory), Yi Yu (Southeast University), Qingyun Li (Harbin Institute of Technology), Junchi Yan (Shanghai Jiao Tong University, Shanghai AI Laboratory), and Yansheng Li (Wuhan University)

Scene-adaptive and Region-aware Multi-modal Prompt for Open Vocabulary Object Detection .. 16741
Xiaowei Zhao (Beihang University,China), Xianglong Liu (Beihang University,China), Duorui Wang (Beihang University,China), Yajun Gao (Beihang University,China), and Zhide Liu (Beihang University, China)

Revisiting the Domain Shift and Sample Uncertainty in Multi-source Active Domain Transfer ... 16751
Wenqiao Zhang (Zhejiang University), Zheqi Lv (Zhejiang University), Hao Zhou (Harbin Institute of Technology), Jia-Wei Liu (National University of Singapore), Juncheng Li (Zhejiang University), Mengze Li (Zhejiang University), Yunfei Li (Ant Group), Dongping Zhang (China Jiliang University), Yueting Zhuang (Zhejiang University), and Siliang Tang (Zhejiang University)

Hyperbolic Learning with Synthetic Captions for Open-World Detection 16762
Fanjie Kong (Duke University), Yanbei Chen (Amazon), Jiarui Cai (Amazon), and Davide Modolo (Amazon)

CricaVPR: Cross-image Correlation-aware Representation Learning for Visual Place Recognition 16772
Feng Lu (Tsinghua University), Xiangyuan Lan (Peng Cheng Laboratory), Lijun Zhang (University of Chinese Academy of Sciences), Dongmei Jiang (Peng Cheng Laboratory), Yaowei Wang (Peng Cheng Laboratory), and Chun Yuan (Tsinghua University)

Point2RBox: Combine Knowledge from Synthetic Visual Patterns for End-to-end Oriented Object Detection with Single Point Supervision 16783
Yi Yu (Southeast University, China), Xue Yang (Shanghai AI Laboratory, China), Qingyun Li (Harbin Institute of Technology, China), Feipeng Da (Southeast University, China), Jifeng Dai (Tsinghua University, China), Yu Qiao (Shanghai AI Laboratory, China), and Junchi Yan (Shanghai Jiao Tong University)

Scene Adaptive Sparse Transformer for Event-based Object Detection 16794
Yansong Peng (University of Science and Technology of China), Hebei Li (University of Science and Technology of China), Yueyi Zhang (University of Science and Technology of China), Xiaoyan Sun (University of Science and Technology of China; Institute of Artificial Intelligence, Hefei Comprehensive National Science Center), and Feng Wu (University of Science and Technology of China; Institute of Artificial Intelligence, Hefei Comprehensive National Science Center)

Visual Delta Generator with Large Multi-modal Models for Semi-supervised Composed Image Retrieval 16805
Young Kyun Jang (Meta AI), Donghyun Kim (Korea University), Zihang Meng (Meta AI), Dat Huynh (Meta AI), and Ser-Nam Lim (University of Central Florida)

Preserving Fairness Generalization in Deepfake Detection 16815
Li Lin (Purdue University, USA), Xinan He (Nanchang University, China), Yan Ju (University at Buffalo, State University of New York, USA), Xin Wang (University at Albany, State University of New York, USA), Feng Ding (Nanchang University, China), and Shu Hu (Purdue University, USA)

Text-to-Image Diffusion Models are Great Sketch-Photo Matchmakers 16826
Subhadeep Koley (University of Surrey, United Kingdom; iFlyTek-Surrey Joint Research Centre on Artificial Intelligence), Ayan Kumar Bhunia (University of Surrey, United Kingdom), Aneeshan Sain (University of Surrey, United Kingdom), Pinaki Nath Chowdhury (University of Surrey, United Kingdom), Tao Xiang (University of Surrey, United Kingdom; iFlyTek-Surrey Joint Research Centre on Artificial Intelligence), and Yi-Zhe Song (University of Surrey, United Kingdom; iFlyTek-Surrey Joint Research Centre on Artificial Intelligence)

Structured Model Probing: Empowering Efficient Transfer Learning by Structured Regularization 16838
Zhi-Fan Wu (Alibaba Group), Chaojie Mao (Alibaba Group), Xue Wang (Alibaba Group), Jianwen Jiang (Alibaba Group), Yiliang Lv (Alibaba Group), and Rong Jin (Alibaba Group)

PromptAD: Learning Prompts with only Normal Samples for Few-Shot Anomaly Detection 16848
Xiaofan Li (East China Normal University), Zhizhong Zhang (East China Normal University), Xin Tan (East China Normal University; Chongqing Institute of East China Normal University), Chengwei Chen (The Navy Military Medical University), Yanyun Qu (Xiamen University), Yuan Xie (East China Normal University; Chongqing Institute of East China Normal University), and Lizhuang Ma (East China Normal University)

How to Handle Sketch-Abstraction in Sketch-Based Image Retrieval? 16859
Subhadeep Koley (University of Surrey, United Kingdom; iFlyTek-Surrey Joint Research Centre on Artificial Intelligence), Ayan Kumar Bhunia (University of Surrey, United Kingdom), Aneeshan Sain (University of Surrey, United Kingdom), Pinaki Nath Chowdhury (University of Surrey, United Kingdom), Tao Xiang (University of Surrey, United Kingdom; iFlyTek-Surrey Joint Research Centre on Artificial Intelligence), and Yi-Zhe Song (University of Surrey, United Kingdom; iFlyTek-Surrey Joint Research Centre on Artificial Intelligence)

Shallow-Deep Collaborative Learning for Unsupervised Visible-Infrared Person Re-Identification 16870
Bin Yang (Wuhan University, China), Jun Chen (Wuhan University, China), and Mang Ye (Wuhan University, China)

Solving the Catastrophic Forgetting Problem in Generalized Category Discovery 16880
Xinzi Cao (Sun Yat-sen University, Peng Cheng Laboratory), Xiawu Zheng (Xiamen University, Peng Cheng Laboratory), Guanhong Wang (Zhejiang University), Weijiang Yu (Sun Yat-sen University), Yunhang Shen (Tecent Youtu Lab), Ke Li (Tecent Youtu Lab), Yutong Lu (Sun Yat-sen University), and Yonghong Tian (Peng Cheng Laboratory, Peking University)

Active Generalized Category Discovery 16890
Shijie Ma (MAIS, Institute of Automation, Chinese Academy of Sciences, China), Fei Zhu (Centre for Artificial Intelligence and Robotics, HKISI-CAS, China), Zhun Zhong (University of Nottingham, NG8 1BB Nottingham, UK), Xu-Yao Zhang (MAIS, Institute of Automation, Chinese Academy of Sciences, China), and Cheng-Lin Liu (MAIS, Institute of Automation, Chinese Academy of Sciences, China)

YOLO-World: Real-Time Open-Vocabulary Object Detection 16901
Tianheng Cheng (Huazhong University of Science and Technology, China), Lin Song (Tencent AI Lab, China), Yixiao Ge (Tencent AI Lab, PCG ARC Lab, Tencent, China), Wenyu Liu (Huazhong University of Science and Technology, China), Xinggang Wang (Huazhong University of Science and Technology, China), and Ying Shan (Tencent AI Lab, PCG ARC Lab, Tencent, China)

Theoretically Achieving Continuous Representation of Oriented Bounding Boxes 16912
Zikai Xiao (BNRist, Tsinghua University), Guoye Yang (BNRist, Tsinghua University), Xue Yang (Shanghai Jiao Tong University), Taijiang Mu (BNRist, Tsinghua University), Junchi Yan (Shanghai Jiao Tong University), and Shimin Hu (BNRist, Tsinghua University)

Decoupled Pseudo-labeling for Semi-Supervised Monocular 3D Object Detection 16923
Jiacheng Zhang (Sun Yat-sen University), Jiaming Li (Sun Yat-sen University), Xiangru Lin (Baidu Inc.), Wei Zhang (Baidu Inc.), Xiao Tan (Baidu Inc.), Junyu Han (Baidu Inc.), Errui Ding (Baidu Inc.), Jingdong Wang (Baidu Inc.), and Guanbin Li (Sun Yat-sen University)

LEOD: Label-Efficient Object Detection for Event Cameras 16933
Ziyi Wu (University of Toronto, Canada), Mathias Gehrig (University of Zurich, Switzerland), Qing Lyu (University of Toronto, Canada), Xudong Liu (University of Toronto, Canada), and Igor Gilitschenski (University of Toronto, Canada)

Lane2Seq: Towards Unified Lane Detection via Sequence Generation 16944
Kunyang Zhou (Southeast University, China)

Open-World Human-Object Interaction Detection via Multi-modal Prompts 16954
Jie Yang (The Chinese University of Hong Kong, Shenzhen), Bingliang Li (The Chinese University of Hong Kong, Shenzhen), Ailing Zeng (International Digital Economy Academy), Lei Zhang (International Digital Economy Academy), and Ruimao Zhang (The Chinese University of Hong Kong, Shenzhen)

DETRs Beat YOLOs on Real-time Object Detection 16965
Yian Zhao (Peking University, China), Wenyu Lv (Baidu, China), Shangliang Xu (Baidu, China), Jinman Wei (Baidu, China), Guanzhong Wang (Baidu, China), Qingqing Dang (Baidu, China), Yi Liu (Baidu, China), and Jie Chen (Peking University, China)

Exploring Region-Word Alignment in Built-in Detector for Open-Vocabulary Object Detection .. 16975
Heng Zhang (JD.com, China), Qiuyu Zhao (JD.com, China), Linyu Zheng (JD.com, China), Hao Zeng (JD.com, China), Zhiwei Ge (JD.com, China), Tianhao Li (JD.com, China), and Sulong Xu (JD.com, China)

Referring Expression Counting 16985
Siyang Dai (Singapore University of Technology and Design), Jun Liu (Singapore University of Technology and Design), and Ngai-Man Cheung (Singapore University of Technology and Design)

ActiveDC: Distribution Calibration for Active Finetuning 16996
Wenshuai Xu (Beihang University, China), Zhenghui Hu (Beihang University, China), Yu Lu (Beihang University, China), Jinzhou Meng (Beihang University, China), Qingjie Liu (Beihang University, China), and Yunhong Wang (Beihang University, China)

LaRE^2: Latent Reconstruction Error Based Method for Diffusion-Generated Image Detection ... 17006
Yunpeng Luo (Tencent YouTu Lab), Junlong Du (Tencent YouTu Lab), Ke Yan (Tencent YouTu Lab), and Shouhong Ding (Tencent YouTu Lab)

Fine-grained Prototypical Voting with Heterogeneous Mixup for Semi-supervised 2D-3D Cross-modal Retrieval 17016
Fan Zhang (Tianjin University, China), Xian-Sheng Hua (Terminus Group, China), Chong Chen (Terminus Group, China), and Xiao Luo (University of California, Los Angeles, USA)

MS-DETR: Efficient DETR Training with Mixed Supervision 17027
Chuyang Zhao (Baidu VIS, China), Yifan Sun (Baidu VIS, China), Wenhao Wang (University of Technology Sydney, Australia), Qiang Chen (Baidu VIS, China), Errui Ding (Baidu VIS, China), Yi Yang (Zhejiang University, China), and Jingdong Wang (Baidu VIS, China)

Context-based and Diversity-driven Specificity in Compositional Zero-Shot Learning 17037
Yun Li (CSIRO's Data61, Australia), Zhe Liu (Bytedance Ltd, Australia), Hang Chen (Snap Inc.), and Lina Yao (CSIRO's Data61, Australia)

Pixel-level Semantic Correspondence through Layout-aware Representation Learning and Multi-scale Matching Integration 17047
Yixuan Sun (Fudan University, China), Zhangyue Yin (Fudan University, China), Haibo Wang (Fudan University, China), Yan Wang (Fudan University, China), Xipeng Qiu (Fudan University, China), Weifeng Ge (Fudan University, China), and Wenqiang Zhang (Fudan University, China)

Exploiting Inter-sample and Inter-feature Relations in Dataset Distillation 17057
Wenxiao Deng (State Key Laboratory for Novel Software Technology, Nanjing University, China), Wenbin Li (State Key Laboratory for Novel Software Technology, Nanjing University, China), Tianyu Ding (Microsoft Corporation, USA), Lei Wang (University of Wollongong, Australia), Hongguang Zhang (Systems Engineering Institute, AMS, China), Kuihua Huang (University of Defense Technology), Jing Huo (State Key Laboratory for Novel Software Technology, Nanjing University, China), and Yang Gao (State Key Laboratory for Novel Software Technology, Nanjing University, China)

Point, Segment and Count: A Generalized Framework for Object Counting 17067
Zhizhong Huang (Shanghai Key Lab of Intelligent Information Processing, Fudan University), Mingliang Dai (Shanghai Key Lab of Intelligent Information Processing, Fudan University), Yi Zhang (Sichuan University), Junping Zhang (Shanghai Key Lab of Intelligent Information Processing, Fudan University), and Hongming Shan (Fudan University)

Dual Pose-invariant Embeddings: Learning Category and Object-specific Discriminative Representations for Recognition and Retrieval 17077
Rohan Sarkar (Purdue University, USA) and Avinash Kak (Purdue University, USA)

Riemannian Multinomial Logistics Regression for SPD Neural Networks 17086
Ziheng Chen (University of Trento, Italy), Yue Song (University of Trento, Italy), Gaowen Liu (Cisco Systems, USA), Ramana Rao Kompella (Cisco Systems, USA), Xiao-Jun Wu (Jiangnan University, China), and Nicu Sebe (University of Trento, Italy)

Learning for Transductive Threshold Calibration in Open-World Recognition 17097
Qin Zhang (Amazon Inc), Dongsheng An (Amazon Inc), Tianjun Xiao (Amazon Inc), Tong He (Amazon Inc), Qingming Tang (Amazon Inc), Ying Nian Wu (Amazon Inc), Joseph Tighe (Amazon Inc), and Yifan Xing (Amazon Inc)

Region-Based Representations Revisited 17107
Michal Shlapentokh-Rothman (University of Illinois at Urbana-Champaign), Ansel Blume (University of Illinois at Urbana-Champaign), Yao Xiao (University of Illinois at Urbana-Champaign), Yuqun Wu (University of Illinois at Urbana-Champaign), Sethuraman TV (University of Illinois at Urbana-Champaign), Heyi Tao (University of Illinois at Urbana-Champaign), Jae Yong Lee (University of Illinois at Urbana-Champaign), Wilfredo Torres (Reconstruct), Yu-Xiong Wang (University of Illinois at Urbana-Champaign), and Derek Hoiem (University of Illinois at Urbana-Champaign)

Magic Tokens: Select Diverse Tokens for Multi-modal Object Re-Identification 17117
Pingping Zhang (Dalian University of Technology, China; Anhui University, China), Yuhao Wang (Dalian University of Technology, China), Yang Liu (Dalian University of Technology, China), Zhengzheng Tu (Anhui University, China), and Huchuan Lu (Dalian University of Technology, China)

Harnessing the Power of MLLMs for Transferable Text-to-Image Person ReID 17127
Wentan Tan (South China University of Technology, China; JD Explore Academy, China), Changxing Ding (South China University of Technology, China; Pazhou Lab, China), Jiayu Jiang (South China University of Technology, China), Fei Wang (South China University of Technology, China;JD Explore Academy, China), Yibing Zhan (JD Explore Academy, China), and Dapeng Tao (Yunnan University, China; Yunnan United Vision Technology Co., Ltd., China)

Holistic Features are almost Sufficient for Text-to-Video Retrieval 17138
Kaibin Tian (MoE Key Lab of DEKE, Renmin University of China), Ruixiang Zhao (MoE Key Lab of DEKE, Renmin University of China), Zijie Xin (MoE Key Lab of DEKE, Renmin University of China; Sichuan University), Bangxiang Lan (MoE Key Lab of DEKE, Renmin University of China), and Xirong Li (MoE Key Lab of DEKE, Renmin University of China)

Enhancing the Power of OOD Detection via Sample-Aware Model Selection 17148
Feng Xue (Shanghai Jiao Tong University, China), Zi He (Hunan University, China), Yuan Zhang (Beijing Normal University, China), Chuanlong Xie (Beijing Normal University, China), Zhenguo Li (Huawei Noah's Ark Lab, China), and Falong Tan (Hunan University, China)

PanoOcc: Unified Occupancy Representation for Camera-based 3D Panoptic Segmentation 17158
Yuqi Wang (Institute of Automation, Chinese Academy of Sciences, China), Yuntao Chen (Centre for Artificial Intelligence and Robotics (HKISI_CAS), China), Xingyu Liao (Independent Researcher), Lue Fan (Institute of Automation, Chinese Academy of Sciences, China), and Zhaoxiang Zhang (Institute of Automation, Chinese Academy of Sciences, China)

VSCode: General Visual Salient and Camouflaged Object Detection with 2D Prompt Learning ... 17169
Ziyang Luo (Northwestern Polytechnical University), Nian Liu (Mohamed bin Zayed University of Artificial Intelligence), Wangbo Zhao (National University of Singapore), Xuguang Yang (Northwestern Polytechnical University), Dingwen Zhang (Northwestern Polytechnical University), Deng-Ping Fan (Nankai International Advanced Research Institute (SHENZHEN FUTIAN) & CS, Nankai University), Fahad Khan (Mohamed bin Zayed University of Artificial Intelligence; CVL, Linköping University), and Junwei Han (Northwestern Polytechnical University)

D3still: Decoupled Differential Distillation for Asymmetric Image Retrieval 17181
Yi Xie (South China University of Technology), Yihong Lin (South China University of Technology), Wenjie Cai (Independent Researcher), Xuemiao Xu (South China University of Technology, State Key Laboratory of Subtropical Building Science, Guangdong Provincial Key Lab of Computational Intelligence and Cyberspace Information, Ministry of Education Key Laboratory of Big Data and Intelligent Robot), Huaidong Zhang (South China University of Technology), Yong Du (Ocean University of China), and Shengfeng He (Singapore Management University)

SFOD: Spiking Fusion Object Detector 17191
Yimeng Fan (Tianjin University, China), Wei Zhang (Tianjin University, China), Changsong Liu (Tianjin University, China), Mingyang Li (Tianjin University, China), and Wenrui Lu (Tianjin University, China)

Depth-Aware Concealed Crop Detection in Dense Agricultural Scenes 17201
Liqiong Wang (China Three Gorges University, China), Jinyu Yang (Southern University of Science and Technology, China; University of Birmingham, U.K.), Yanfu Zhang (College of William and Mary, USA), Fangyi Wang (China Three Gorges University, China), and Feng Zheng (Southern University of Science and Technology, China)

Extreme Point Supervised Instance Segmentation 17212
Hyeonjun Lee (Lunit Inc, South Korea), Sehyun Hwang (Postech, South Korea), and Suha Kwak (Postech, South Korea)

Enhance Image Classification via Inter-Class Image Mixup with Diffusion Model 17223
Zhicai Wang (University of Science and Technology of China), Longhui Wei (Huawei Inc.), Tan Wang (Nanyang Technological University), Heyu Chen (University of Science and Technology of China), Yanbin Hao (University of Science and Technology of China), Xiang Wang (University of Science and Technology of China), Xiangnan He (University of Science and Technology of China), and Qi Tian (Huawei Inc.)

Multimodal Industrial Anomaly Detection by Crossmodal Feature Mapping 17234
Alex Costanzino (University of Bologna, Italy), Pierluigi Zama Ramirez (University of Bologna, Italy), Giuseppe Lisanti (University of Bologna, Italy), and Luigi Di Stefano (University of Bologna, Italy)

Effective Video Mirror Detection with Inconsistent Motion Cues 17244
Alex Warren (Swansea University, United Kingdom), Ke Xu (City University of Hong Kong, Hong Kong S.A.R China), Jiaying Lin (City University of Hong Kong, Hong Kong S.A.R China), Gary K.L. Tam (Swansea University, United Kingdom), and Rynson W.H. Lau (City University of Hong Kong, Hong Kong S.A.R China)

Multi-Attribute Interactions Matter for 3D Visual Grounding 17253
Can Xu (Nanjing University of Science and Technology), Yuehui Han (Nanjing University of Science and Technology), Rui Xu (Nanjing University of Science and Technology), Le Hui (Northwestern Polytechnical University), Jin Xie (Nanjing University of Science and Technology), and Jian Yang (Nanjing University of Science and Technology)

Looking 3D: Anomaly Detection with 2D-3D Alignment 17263
Ankan Bhunia (University of Edinburgh), Changjian Li (University of Edinburgh), and Hakan Bilen (University of Edinburgh)

Characteristics Matching Based Hash Codes Generation for Efficient Fine-grained Image Retrieval 17273
Zhen-Duo Chen (Shandong University, China), Li-Jun Zhao (Shandong University, China), Zi-Chao Zhang (Shandong University, China), Xin Luo (Shandong University, China), and Xin-Shun Xu (Shandong University, China)

EASE-DETR: Easing the Competition among Object Queries 17282
Yulu Gao (Beihang University, China), Yifan Sun (Baidu VIS, China), Xudong Ding (Beihang University, China), Chuyang Zhao (Baidu VIS, China), and Si Liu (Beihang University, China)

ProS: Prompting-to-simulate Generalized knowledge for Universal Cross-Domain Retrieval 17292
Kaipeng Fang (University of Electronic Science and Technology of China), Jingkuan Song (University of Electronic Science and Technology of China), Lianli Gao (University of Electronic Science and Technology of China), Pengpeng Zeng (University of Electronic Science and Technology of China), Zhi-Qi Cheng (Carnegie Mellon University), Xiyao Li (Kuaishou Technology), and Heng Tao Shen (Tongji University)

Exploring Orthogonality in Open World Object Detection 17302
Zhicheng Sun (Peking University, China), Jinghan Li (Peking University, China), and Yadong Mu (Peking University, China)

A Generative Approach for Wikipedia-Scale Visual Entity Recognition 17313
Mathilde Caron (Google Research, France), Ahmet Iscen (Google Research, France), Alireza Fathi (Google Research), and Cordelia Schmid (Google Research)

Unleashing Channel Potential: Space-Frequency Selection Convolution for SAR Object Detection 17323
Ke Li (Xidian University, China), Di Wang (Xidian University, China), Zhangyuan Hu (Xidian University, China), Wenxuan Zhu (Xidian University, China), Shaofeng Li (Xidian University, China), and Quan Wang (Xidian University, China)

Hyperspherical Classification with Dynamic Label-to-Prototype Assignment 17333
Mohammad Saeed Ebrahimi Saadabadi (West Virginia University, USA), Ali Dabouei (Carnegie Mellon University, USA), Sahar Rahimi Malakshan (West Virginia University, USA), and Nasser M. Nasrabadi (West Virginia University, USA)

A Pedestrian is Worth One Prompt: Towards Language Guidance Person Re-Identification 17343
Zexian Yang (Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), Dayan Wu (Chinese Academy of Sciences, China), Chenming Wu (Baidu Inc, China), Zheng Lin (Chinese Academy of Sciences, China), Jingzi Gu (Chinese Academy of Sciences, China), and Weiping Wang (Chinese Academy of Sciences, China)

VSRD: Instance-Aware Volumetric Silhouette Rendering for Weakly Supervised 3D Object Detection 17354
Zihua Liu (Tokyo Institute of Technology, Japan), Hiroki Sakuma (T2 Inc., Japan), and Masatoshi Okutomi (Tokyo Institute of Technology, Japan)

Improving Visual Recognition with Hyperbolical Visual Hierarchy Mapping 17364
Hyeongjun Kwon (Yonsei University), Jinhyun Jang (Yonsei University), Jin Kim (Yonsei University), Kwonyoung Kim (Yonsei University), and Kwanghoon Sohn (Yonsei University)

On Train-Test Class Overlap and Detection for Image Retrieval 17375
Chull Hwan Song (Dealicious Inc.), Jooyoung Yoon (Dealicious Inc.), Taebaek Hwang (Dealicious Inc.), Shunghyun Choi (Dealicious Inc.), Yeong Hyeon Gu (Sejong University), and Yannis Avrithis (Institute of Advanced Research on Artificial Intelligence (IARAI))

Multi-Scale Video Anomaly Detection by Multi-Grained Spatio-Temporal Representation Learning 17385
Menghao Zhang (State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications), Jingyu Wang (State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications), Qi Qi (State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications), Haifeng Sun (State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications), Zirui Zhuang (State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications), Pengfei Ren (State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications), Ruilong Ma (State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications), and Jianxin Liao (State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications)

LAA-Net: Localized Artifact Attention Network for Quality-Agnostic and Generalizable Deepfake Detection 17395
Dat Nguyen (University of Luxembourg, Luxembourg), Nesryne Mejri (University of Luxembourg, Luxembourg), Inder Pal Singh (University of Luxembourg, Luxembourg), Polina Kuleshova (University of Luxembourg, Luxembourg), Marcella Astrid (University of Luxembourg, Luxembourg), Anis Kacem (University of Luxembourg, Luxembourg), Enjie Ghorbel (University of Manouba, Tunisia), and Djamila Aouada (University of Luxembourg, Luxembourg)

Rethinking Boundary Discontinuity Problem for Oriented Object Detection 17406
Hang Xu (Hangzhou Dianzi University, China), Xinyuan Liu (Institute of Computing Technology, Chinese Academy of Sciences, China), Haonan Xu (Institute of Computing Technology, Chinese Academy of Sciences, China), Yike Ma (Institute of Computing Technology, Chinese Academy of Sciences, China), Zunjie Zhu (Hangzhou Dianzi University, China), Chenggang Yan (Hangzhou Dianzi University, China), and Feng Dai (Institute of Computing Technology, Chinese Academy of Sciences, China)

Hybrid Proposal Refiner: Revisiting DETR Series from the Faster R-CNN Perspective 17416
Jinjing Zhao (The University of Sydney), Fangyun Wei (The University of Sydney), and Chang Xu (The University of Sydney)

Retrieval-Augmented Open-Vocabulary Object Detection 17427
Jooyeon Kim (Korea University, South Korea), Eulrang Cho (Samsung Research, South Korea), Sehyung Kim (Korea University, South Korea), and Hyunwoo J. Kim (Korea University, South Korea)

LiDAR-based Person Re-identification 17437
Wenxuan Guo (Tsinghua University, China), Zhiyu Pan (Tsinghua University, China), Yingping Liang (Beijing Institute of Technology, China), Ziheng Xi (Tsinghua University, China), Zhicheng Zhong (Tsinghua University, China), Jianjiang Feng (Tsinghua University, China), and Jie Zhou (Tsinghua University, China)

EventDance: Unsupervised Source-free Cross-modal Adaptation for Event-based Object Recognition 17448
Xu Zheng (The Hong Kong University of Science and Technology (Guangzhou), China) and Lin Wang (The Hong Kong University of Science and Technology (Guangzhou), China)

All in One Framework for Multimodal Re-identification in the Wild 17459
He Li (Wuhan University), Mang Ye (Wuhan University), Ming Zhang (Guangzhou Urban Planning Design Survey Research Institute), and Bo Du (Wuhan University)

Logarithmic Lenses: Exploring Log RGB Data for Image Classification 17470
Bruce A. Maxwell (Northeastern University), Sumegha Singhania (Northeastern University), Avnish Patel (Northeastern University), Rahul Kumar (Northeastern University), Heather Fryling (Northeastern University), Sihan Li (Northeastern University), Haonan Sun (Northeastern University), Ping He (Northeastern University), and Zewen Li (Northeastern University)

ID-like Prompt Learning for Few-Shot Out-of-Distribution Detection 17480
Yichen Bai (Tianjin University), Zongbo Han (Tianjin University), Bing Cao (Tianjin University; Tianjin Key Lab of Machine Learning), Xiaoheng Jiang (Zhengzhou University), Qinghua Hu (Tianjin University;Tianjin Key Lab of Machine Learning), and Changqing Zhang (Tianjin University;Tianjin Key Lab of Machine Learning)

Infrared Small Target Detection with Scale and Location Sensitivity 17490
Qiankun Liu (Beijing Institute of Technology, China), Rui Liu (Beijing Institute of Technology, China), Bolun Zheng (Hangzhou Dianzi University, China), Hongkui Wang (Hangzhou Dianzi University, China), and Ying Fu (Beijing Institute of Technology)

SURE: SUrvey REcipes for building reliable and robust deep networks 17500
Yuting Li (Intellindust, China; China Three Gorges University, China), Yingyi Chen (KU Leuven, Belgium), Xuanlong Yu (Paris-Saclay University, France; ENSTA - Institut Polytechnique de Paris, France), Dexiong Chen (Max Planck Institute of Biochemistry, Germany), and Xi Shen (Intellindust, China)

Hyperbolic Anomaly Detection 17511
Huimin Li (Beihang University, China), Zhentao Chen (Beihang University, China), Yunhao Xu (Beihang University, China), and Junlin Hu (Beihang University, China)

Instruct-ReID: A Multi-purpose Person Re-identification Task with Instructions 17521
Weizhen He (Zhejiang University, China), Yiheng Deng (Zhejiang University, China), Shixiang Tang (The University of Sydney, Australia), Qihao Chen (Liaoning Technical University, China), Qingsong Xie (Shanghai Jiao Tong University, China), Yizhou Wang (Shanghai AI Laboratory, China), Lei Bai (Shanghai AI Laboratory, China), Feng Zhu (SenseTime Research, China), Rui Zhao (SenseTime Research, China), Wanli Ouyang (Shanghai AI Laboratory, China), Donglian Qi (Zhejiang University, China), and Yunfeng Yan (Zhejiang University)

CA-Jaccard: Camera-aware Jaccard Distance for Person Re-identification 17532
Yiyu Chen (Beijing Institute of Technology, China), Zheyi Fan (Beijing Institute of Technology, China), Zhaoru Chen (Beijing Institute of Technology, China), and Yixuan Zhu (Beijing Institute of Technology, China)

Improved Zero-Shot Classification by Adapting VLMs with Text Descriptions 17542
Oindrila Saha (UMass Amherst), Grant Van Horn (UMass Amherst), and Subhransu Maji (UMass Amherst)

Modeling Collaborator: Enabling Subjective Vision Classification With Minimal Human Effort via LLM Tool-Use 17553
Imad Eddine Toubal (University of Missouri), Aditya Avinash (Google Research), Neil Gordon Alldrin (Google Research), Jan Dlabal (Google Research), Wenlei Zhou (Google Research), Enming Luo (Google Research), Otilia Stretcu (Google Research), Hao Xiong (Google Research), Chun-Ta Lu (Google Research), Howard Zhou (Google Research), Ranjay Krishna (University of Washington), Ariel Fuxman (Google Research), and Tom Duerig (Google Research)

Neural Exposure Fusion for High-Dynamic Range Object Detection 17564
Emmanuel Onzon (Torc Robotics, Germany), Maximilian Bömer (Torc Robotics, Germany), Fahim Mannan (Torc Robotics, Canada), and Felix Heide (Princeton University, USA)

Salience DETR: Enhancing Detection Transformer with Hierarchical Salience Filtering Refinement 17574
Xiuquan Hou (Xi'an Jiaotong University, China), Meiqin Liu (Xi'an Jiaotong University, China), Senlin Zhang (Zhejiang University, China), Ping Wei (Xi'an Jiaotong University, China), and Badong Chen (Xi'an Jiaotong University, China)

Learning Transferable Negative Prompts for Out-of-Distribution Detection 17584
Tianqi Li (Beihang University, China), Guansong Pang (Singapore Management University, Singapore), Xiao Bai (Beihang University, China), Wenjun Miao (Beihang University), and Jin Zheng (Beihang University)

TransLoc4D: Transformer-based 4D Radar Place Recognition 17595
Guohao Peng (Nanyang Technological University), Heshan Li (Nanyang Technological University), Yangyang Zhao (Nanyang Technological University), Jun Zhang (Nanyang Technological University), Zhenyu Wu (Nanyang Technological University), Pengyu Zheng (Nanyang Technological University), and Danwei Wang (Nanyang Technological University)

Prompt-Driven Dynamic Object-Centric Learning for Single Domain Generalization 17606
Deng Li (Tianjin University, China), Aming Wu (Xidian University, China), Yaowei Wang (Peng Cheng Laboratory, China), and Yahong Han (Tianjin University, China)

Anomaly Heterogeneity Learning for Open-set Supervised Anomaly Detection 17616
Jiawen Zhu (Singapore Management University), Choubo Ding (University of Adelaide), Yu Tian (Harvard University), and Guansong Pang (Singapore Management University)

Contrastive Learning for DeepFake Classification and Localization via Multi-Label Ranking 17627
Cheng-Yao Hong (Institute of Information Science, Academia Sinica, Taiwan), Yen-Chi Hsu (Institute of Information Science, Academia Sinica, Taiwan), and Tyng-Luh Liu (Institute of Information Science, Academia Sinica, Taiwan)

Adaptive Softassign via Hadamard-Equipped Sinkhorn 17638
Binrui Shen (Xi'an Jiaotong-Liverpool University, ChinaUniversity of Liverpool, Liverpool, United Kingdom), Qiang Niu (Xi'an Jiaotong-Liverpool University, China), and Shengxin Zhu (Beijing Normal University, China; BNU-HKBU United International College, China)

An Asymmetric Augmented Self-Supervised Learning Method for Unsupervised Fine-Grained Image Hashing 17648
Feiran Hu (Nanjing University of Science and Technology, China), Chenlin Zhang (4Paradigm Inc., China), Jiangliang Guo (AInnovation Technology Group Co., Ltd, China), Xiu-Shen Wei (Southeast University, China), Lin Zhao (Nanjing University of Science and Technology, China), Anqi Xu (University of Toronto, Canada), and Lingyan Gao (AInnovation Technology Group Co., Ltd, China)

Optimal Transport Aggregation for Visual Place Recognition 17658
Sergio Izquierdo (I3A - University of Zaragoza) and Javier Civera (I3A - University of Zaragoza)

Atom-Level Optical Chemical Structure Recognition with Limited Supervision 17669
Martijn Oldenhof (KU Leuven, Belgium), Edward De Brouwer (KU Leuven, Belgium), Adam Arany (KU Leuven, Belgium), and Yves Moreau (KU Leuven, Belgium)

Novel Class Discovery for Ultra-Fine-Grained Visual Categorization 17679
Yu Liu (Dalian University of Technology, China), Yaqi Cai (Dalian University of Technology, China), Qi Jia (Dalian University of Technology, China), Binglin Qiu (Dalian University of Technology, China), Weimin Wang (Dalian University of Technology, China), and Nan Pu (University of Trento, Italy)

Attribute-Guided Pedestrian Retrieval: Bridging Person Re-ID with Internal Attribute Variability 17689
Yan Huang (Center for Research on Intelligent Perception and Computing, State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China), Zhang Zhang (Center for Research on Intelligent Perception and Computing, State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China; University of Chinese Academy of Sciences), Qiang Wu (University of Technology Sydney, Australia), Yi Zhong (Beijing Institute of Technology, China), and Liang Wang (Center for Research on Intelligent Perception and Computing, State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China; University of Chinese Academy of Sciences)

Robust Noisy Correspondence Learning with Equivariant Similarity Consistency 17700
Yuchen Yang (Xidian University, China), Likai Wang (Xidian University, China), Erkun Yang (Xidian University, China), and Cheng Deng (Xidian University, China)

Bootstrapping SparseFormers from Vision Foundation Models 17710
Ziteng Gao (National University of Singapore, Singapore), Zhan Tong (National University of Singapore, Singapore), Kevin Qinghong Lin (National University of Singapore, Singapore), Joya Chen (National University of Singapore, Singapore), and Mike Zheng Shou (National University of Singapore, Singapore)

Not All Classes Stand on Same Embeddings: Calibrating a Semantic Distance with Metric Tensor 17722
Jae Hyeon Park (Dongguk University, Republic of Korea), Gyoomin Lee (Dongguk University, Republic of Korea), Seunggi Park (Dongguk University, Republic of Korea), and Sung In Cho (Dongguk University, Republic of Korea)

Improving Single Domain-Generalized Object Detection: A Focus on Diversification and Alignment 17732
Muhammad Sohail Danish (Mohamed bin Zayed University of Artificial Intelligence, UAE), Muhammad Haris Khan (Mohamed bin Zayed University of Artificial Intelligence, UAE), Muhammad Akhtar Munir (Mohamed bin Zayed University of Artificial Intelligence, UAE), M. Saquib Sarfraz (Karlsruhe Institute of Technology, Germany), and Mohsen Ali (Information Technology University (ITU), Pakistan)

On the Estimation of Image-matching Uncertainty in Visual Place Recognition 17743
Mubariz Zaffar (TU Delft), Liangliang Nan (TU Delft), and Julian F. P. Kooij (TU Delft)

Supervised Anomaly Detection for Complex Industrial Images 17754
Aimira Baitieva (valeo), David Hurych (valeo.ai), Victor Besnier (valeo.ai), and Olivier Bernard (valeo)

Fourier-basis Functions to Bridge Augmentation Gap: Rethinking Frequency Augmentation in Image Classification 17763
Puru Vaish (Unviersity of Twente), Shunxin Wang (University of Twente), and Nicola Strisciuglio (University of Twente)

TransNeXt: Robust Foveal Visual Perception for Vision Transformers 17773
Dai Shi (Independent Researcher)

Plug and Play Active Learning for Object Detection 17784
Chenhongyi Yang (University of Edinburgh, UK), Lichao Huang (Horizon Robotics, China), and Elliot J. Crowley (University of Edinburgh, UK)

BoQ: A Place is Worth a Bag of Learnable Queries 17794
Amar Ali-bey (Université Laval, Québec, Canada), Brahim Chaib-draa (Université Laval, Québec, Canada), and Philippe Giguère (Université Laval, Québec, Canada)

From Coarse to Fine-Grained Open-Set Recognition 17804
Nico Lang (University of Copenhagen, Denmark), Vésteinn Snæbjarnarson (University of Copenhagen, Denmark), Elijah Cole (Altos Labs, USA), Oisin Mac Aodha (University of Edinburgh, Scotland), Christian Igel (University of Copenhagen, Denmark), and Serge Belongie (University of Copenhagen, Denmark)

Exploring Pose-Aware Human-Object Interaction via Hybrid Learning 17815
Eastman Z Y Wu (Tsinghua University, China; Beijing National Research Center for Information Science and Technology (BNRist), China), Yali Li (Tsinghua University, China; Beijing National Research Center for Information Science and Technology (BNRist), China), Yuan Wang (Tsinghua University, China; Beijing National Research Center for Information Science and Technology (BNRist), China), and Shengjin Wang (Tsinghua University, China; Beijing National Research Center for Information Science and Technology (BNRist), China)

Toward Generalist Anomaly Detection via In-context Residual Learning with Few-shot Sample Prompts 17826
Jiawen Zhu (Singapore Management University) and Guansong Pang (Singapore Management University)

Learning to Navigate Efficiently and Precisely in Real Environments 17837
Guillaume Bono (Naver Labs Europe), Hervé Poirier (Naver Labs Europe), Leonid Antsfeld (Naver Labs Europe), Gianluca Monaci (Naver Labs Europe), Boris Chidlovskii (Naver Labs Europe), and Christian Wolf (Naver Labs Europe)

Task-Conditioned Adaptation of Visual Features in Multi-Task Policy Learning 17847
Pierre Marza (INSA Lyon, France), Laetitia Matignon (Université Claude Bernard Lyon 1, France), Olivier Simonin (INSA Lyon, France), and Christian Wolf (Naver Labs Europe, France)

FastMAC: Stochastic Spectral Sampling of Correspondence Graph 17857
Yifei Zhang (AIR, Tsinghua University & University of Chinese Academy of Sciences), Hao Zhao (AIR, Tsinghua University), Hongyang Li (Shanghai AI Laboratory), and Siheng Chen (Shanghai Jiao Tong University & Shanghai AI Laboratory)

FoundationPose: Unified 6D Pose Estimation and Tracking of Novel Objects 17868
Bowen Wen (NVIDIA, USA), Wei Yang (NVIDIA, USA), Jan Kautz (NVIDIA, USA), and Stan Birchfield (NVIDIA, USA)

CAGE: Controllable Articulation GEneration 17880
Jiayi Liu (Simon Fraser University, Canada), Hou In Ivan Tam (Simon Fraser University, Canada), Ali Mahdavi-Amiri (Simon Fraser University, Canada), and Manolis Savva (Simon Fraser University, Canada)

SingularTrajectory: Universal Trajectory Predictor Using Diffusion Model 17890
Inhwan Bae (Gwangju Institute of Science and Technology, South Korea), Young-Jae Park (Gwangju Institute of Science and Technology, South Korea), and Hae-Gon Jeon (Gwangju Institute of Science and Technology, South Korea)

Language-driven Grasp Detection 17902
An Dinh Vuong (FPT Software - AI Center, Vietnam), Minh Nhat Vu (Automation & Control Institute, TU Wien, Austria), Baoru Huang (Imperial College London, UK), Nghia Nguyen (FPT Software - AI Center, Vietnam), Hieu Le (FPT Software - AI Center, Vietnam), Thieu Vo (Ton Duc Thang University, Vietnam), and Anh Nguyen (University of Liverpool, UK)

MemoNav: Working Memory Model for Visual Navigation 17913
Hongxin Li (University of Chinese Academy of Sciences, China & CASIA, China), Zeyu Wang (University of Chinese Academy of Sciences, China & CASIA, China), Xu Yang (University of Chinese Academy of Sciences, China & CASIA, China), Yuran Yang (Tencent, China), Shuqi Mei (Tencent, China), and Zhaoxiang Zhang (University of Chinese Academy of Sciences, China & CASIA, China & Shanghai AI Lab, China)

NOPE: Novel Object Pose Estimation from a Single Image 17923
Van Nguyen Nguyen (Ecole des Ponts ParisTech, France), Thibault Groueix (Adobe, USA), Georgy Ponimatkin (Ecole des Ponts ParisTech, France), Yinlin Hu (Magic Leap, Switzerland), Renaud Marlet (Ecole des Ponts ParisTech, France), Mathieu Salzmann (EPFL), and Vincent Lepetit (Ecole des Ponts ParisTech)

Dexterous Grasp Transformer 17933
Guo-Hao Xu (Sun Yat-sen University, China), Yi-Lin Wei (Sun Yat-sen University, China), Dian Zheng (Sun Yat-sen University, China), Xiao-Ming Wu (Sun Yat-sen University, China), and Wei-Shi Zheng (Sun Yat-sen University, China)

Versatile Navigation Under Partial Observability via Value-guided Diffusion Policy 17943
Gengyu Zhang (Illinois Institute of Technology, USA), Hao Tang (Carnegie Mellon University, USA), and Yan Yan (Illinois Institute of Technology, USA)

CyberDemo: Augmenting Simulated Human Demonstration for Real-World Dexterous Manipulation . 17952
Jun Wang (University of California San Diego, USA), Yuzhe Qin (University of California San Diego, USA), Kaiming Kuang (University of California San Diego, USA), Yigit Korkmaz (University of Southern California, USA), Akhilan Gurumoorthy (University of California San Diego, USA), Hao Su (University of California San Diego, USA), and Xiaolong Wang (University of California San Diego, USA)

SchurVINS: Schur Complement-Based Lightweight Visual Inertial Navigation System 17964
Yunfei Fan (ByteDance, China), Tianyu Zhao (ByteDance, China), and Guidong Wang (ByteDance, China)

READ: Retrieval-Enhanced Asymmetric Diffusion for Motion Planning 17974
Takeru Oba (Toyota Technological Institute), Matthew Walter (Toyota Technological Institute at Chicago), and Norimichi Ukita (Toyota Technological Institute)

Retrieval-Augmented Embodied Agents 17985
Yichen Zhu (Midea Group), Zhicai Ou (Midea Group), Xiaofeng Mou (Midea Group), and Jian Tang (Midea Group)

Collaborative Semantic Occupancy Prediction with Hybrid Feature Fusion in Connected Automated Vehicles 17996
Rui Song (Fraunhofer IVI; Technical University of Munich), Chenwei Liang (Fraunhofer IVI), Hu Cao (Technical University of Munich), Zhiran Yan (Technische Hochschule Ingolstadt), Walter Zimmer (Technical University of Munich), Markus Gross (Fraunhofer IVI), Andreas Festag (Fraunhofer IVI; Technische Hochschule Ingolstadt), and Alois Knoll (Technical University of Munich)

Diffusion-EDFs: Bi-equivariant Denoising Generative Modeling on SE(3) for Visual Robotic Manipulation 18007
Hyunwoo Ryu (Yonsei University, South Korea), Jiwoo Kim (Yonsei University, South Korea), Hyunseok An (Yonsei University, South Korea), Junwoo Chang (Yonsei University, South Korea), Joohwan Seo (University of California, Berkeley), Taehan Kim (Samsung Research, South Korea), Yubin Kim (Massachusetts Institute of Technology), Chaewon Hwang (Ewha Womans University, South Korea), Jongeun Choi (Yonsei University, South Korea), and Roberto Horowitz (University of California, Berkeley)

Adaptive VIO: Deep Visual-Inertial Odometry with Online Continual Learning 18019
Youqi Pan (Peking University), Wugen Zhou (Peking University), Yingdian Cao (Peking University), and Hongbin Zha (Peking University)

F3Loc: Fusion and Filtering for Floorplan Localization 18029
Changan Chen (ETH Zurich, Switzerland), Rui Wang (Microsoft Mixed Reality & AI Lab Zurich, Switzerland), Christoph Vogel (Microsoft Mixed Reality & AI Lab Zurich, Switzerland), and Marc Pollefeys (ETH Zurich, Switzerland, Microsoft Mixed Reality & AI Lab Zurich, Switzerland)

Gaussian Splatting SLAM 18039
Hidenobu Matsuki (Imperial College London), Riku Murai (Imperial College London), Paul H.J. Kelly (Imperial College London), and Andrew J. Davison (Imperial College London)

SUGAR: Pre-training 3D Visual Representations for Robotics 18049
Shizhe Chen (Inria, France), Ricardo Garcia (Inria, France), Ivan Laptev (Mohamed bin Zayed University of Artificial Intelligence, United Arab Emirates), and Cordelia Schmid (Inria, France)

ManipLLM: Embodied Multimodal Large Language Model for Object-Centric Robotic Manipulation... 18061
Xiaoqi Li (Peking University, China), Mingxu Zhang (Beijing University of Posts and Telecommunications), Yiran Geng (Peking University, China), Haoran Geng (Peking University, China), Yuxing Long (Peking University, China), Yan Shen (Peking University, China), Renrui Zhang (The Chinese University of Hong Kong), Jiaming Liu (Peking University, China), and Hao Dong (Peking University, China)

Open-Vocabulary Object 6D Pose Estimation 18071
Jaime Corsetti (Fondazione Bruno Kessler, University of Trento), Davide Boscaini (Fondazione Bruno Kessler), Changjae Oh (Queen Mary University of London), Andrea Cavallaro (Idiap Research Institute, EPFL), and Fabio Poiesi (Fondazione Bruno Kessler)

Hierarchical Diffusion Policy for Kinematics-Aware Multi-Task Robotic Manipulation 18081
Xiao Ma (Dyson Robot Learning Lab), Sumit Patidar (Dyson Robot Learning Lab), Iain Haughton (Dyson Robot Learning Lab), and Stephen James (Dyson Robot Learning Lab)

Smart Help: Strategic Opponent Modeling for Proactive and Adaptive Robot Assistance in Households 18091
Zhihao Cao (Beijing Institute for General Artificial Intelligence (BIGAI), China; Tsinghua University, China), Zidong Wang (Beijing Institute for General Artificial Intelligence (BIGAI), China; Tsinghua University, China), Siwen Xie (Peking University, China), Anji Liu (University of California, Los Angeles, USA), and Lifeng Fan (Beijing Institute for General Artificial Intelligence (BIGAI), China)

Generalizing 6-DoF Grasp Detection via Domain Prior Knowledge 18102
Haoxiang Ma (Beihang University, China), Modi Shi (Beihang University, China), Boyang Gao (Geometry Robotics, China), and Di Huang (Beihang University, China)

A Simple and Effective Point-based Network for Event Camera 6-DOFs Pose Relocalization 18112
Hongwei Ren (The Hong Kong University of Science and Technology (Guangzhou)), Jiadong Zhu (The Hong Kong University of Science and Technology (Guangzhou)), Yue Zhou (The Hong Kong University of Science and Technology (Guangzhou)), Haotian Fu (The Hong Kong University of Science and Technology (Guangzhou)), Yulong Huang (The Hong Kong University of Science and Technology (Guangzhou)), and Bojun Cheng (The Hong Kong University of Science and Technology (Guangzhou))

Neural Visibility Field for Uncertainty-Driven Active Mapping 18122
Shangjie Xue (Georgia Institute of Technology), Jesse Dill (Georgia Institute of Technology), Pranay Mathur (Georgia Institute of Technology), Frank Dellaert (Georgia Institute of Technology), Panagiotis Tsiotra (Georgia Institute of Technology), and Danfei Xu (Georgia Institute of Technology)

SPIN: Simultaneous Perception, Interaction and Navigation 18133
Shagun Uppal (Carnegie Mellon University, USA), Ananye Agarwal (Carnegie Mellon University, USA), Haoyu Xiong (Carnegie Mellon University, USA), Kenneth Shaw (Carnegie Mellon University, USA), and Deepak Pathak (Carnegie Mellon University, USA)

PredToken: Predicting Unknown Tokens and Beyond with Coarse-to-Fine Iterative Decoding ... 18143
Xuesong Nie (Zhejiang University, China), Haoyuan Jin (Zhejiang University, China), Yunfeng Yan (Zhejiang University, China), Xi Chen (The University of Hong Kong, China), Zhihang Zhu (Zhejiang University, China), and Donglian Qi (Zhejiang University, China)

TIM: A Time Interval Machine for Audio-Visual Action Recognition 18153
Jacob Chalk (University of Bristol, United Kingdom), Jaesung Huh (University of Oxford, United Kingdom), Evangelos Kazakos (Czech Technical University in Prague, Czech Republic), Andrew Zisserman (University of Oxford, United Kingdom), and Dima Damen (University of Bristol, United Kingdom)

AutoAD III: The Prequel – Back to the Pixels 18164
Tengda Han (University of Oxford), Max Bain (University of Oxford), Arsha Nagrani (University of Oxford), Gül Varol (École des Ponts ParisTech), Weidi Xie (Shanghai Jiao Tong University), and Andrew Zisserman (University of Oxford)

FACT: Frame-Action Cross-Attention Temporal Modeling for Efficient Action Segmentation 18175
Zijia Lu (Northeastern University, USA) and Ehsan Elhamifar (Northeastern University, USA)

Progress-Aware Online Action Segmentation for Egocentric Procedural Task Videos 18186
Yuhan Shen (Northeastern University) and Ehsan Elhamifar (Northeastern University)

Video ReCap: Recursive Captioning of Hour-Long Videos 18198
Md Mohaiminul Islam (UNC Chapel Hill), Ngan Ho (UNC Chapel Hill), Xitong Yang (Meta), Tushar Nagarajan (Meta), Lorenzo Torresani (Meta), and Gedas Bertasius (UNC Chapel Hill)

OmniViD: A Generative Framework for Universal Video Understanding 18209
Junke Wang (Fudan University, China), Dongdong Chen (Microsoft Cloud + AI, USA), Chong Luo (Microsoft Research Asia, China), Bo He (University of Maryland, College Park, USA), Lu Yuan (Microsoft Cloud + AI, USA), Zuxuan Wu (Fudan University, China), and Yu-Gang Jiang (Fudan University, China)

MovieChat: From Dense Token to Sparse Memory for Long Video Understanding 18221
Enxin Song (Zhejiang University), Wenhao Chai (University of Washington), Guanhong Wang (Zhejiang University), Yucheng Zhang (Zhejiang University), Haoyang Zhou (Zhejiang University), Feiyang Wu (Zhejiang University), Haozhe Chi (Zhejiang University), Xun Guo (Microsoft Research Asia), Tian Ye (Hong Kong University of Science and Technology (GZ)), Yanting Zhang (Donghua University), Yan Lu (Microsoft Research Asia), Jenq-Neng Hwang (University of Washington), and Gaoang Wang (Zhejiang University)

Learning Group Activity Features Through Person Attribute Prediction 18233
Chihiro Nakatani (TTI-J), Hiroaki Kawashima (University of Hyogo), and Norimichi Ukita (TTI-J)

Streaming Dense Video Captioning 18243
Xingyi Zhou (Google), Anurag Arnab (Google), Shyamal Buch (Google), Shen Yan (Google), Austin Myers (Google), Xuehan Xiong (Google), Arsha Nagrani (Google), and Cordelia Schmid (Google)

Efficient and Effective Weakly-Supervised Action Segmentation via Action-Transition-Aware Boundary Alignment 18253
Angchi Xu (Sun Yat-sen University, China) and Wei-Shi Zheng (Sun Yat-sen University, China)

Benchmarking the Robustness of Temporal Action Detection Models Against Temporal Corruptions 18263
Runhao Zeng (Shenzhen MSU-BIT University), Xiaoyong Chen (Shenzhen University), Jiaming Liang (Shenzhen University), Huisi Wu (Shenzhen University), Guangzhong Cao (Shenzhen University), and Yong Guo (South China University of Technology)

A Backpack Full of Skills: Egocentric Video Understanding with Diverse Task Perspectives 18275
Simone Alberto Peirone (Politecnico di Torino, Italy), Francesca Pistilli (Politecnico di Torino, Italy), Antonio Alliegro (Politecnico di Torino and Istituto Italiano di Tecnologia, Italy), and Giuseppe Averta (Politecnico di Torino, Italy)

Summarize the Past to Predict the Future: Natural Language Descriptions of Context Boost Multimodal Object Interaction Anticipation 18286
Razvan-George Pasca (ETH Zurich), Alexey Gavryushin (ETH Zurich), Muhammad Hamza (University of Zurich), Yen-Ling Kuo (University of Virginia), Kaichun Mo (NVIDIA), Luc Van Gool (ETH Zurich; KU Leuven; INSAIT), Otmar Hilliges (ETH Zurich), and Xi Wang (ETH Zurich)

Open-Vocabulary Video Anomaly Detection 18297
Peng Wu (Northwestern Polytechnical University), Xuerong Zhou (Northwestern Polytechnical University), Guansong Pang (Singapore Management University), Yujia Sun (Xidian University), Jing Liu (Xidian University), Peng Wang (Northwestern Polytechnical University), and Yanning Zhang (Northwestern Polytechnical University)

Task-Driven Exploration: Decoupling and Inter-Task Feedback for Joint Moment Retrieval and Highlight Detection 18308
Jin Yang (Xi'an Jiaotong University, China), Ping Wei (Xi'an Jiaotong University, China), Huan Li (Xi'an Jiaotong University, China), and Ziyang Ren (Xi'an Jiaotong University, China)

Prompt-Enhanced Multiple Instance Learning for Weakly Supervised Video Anomaly Detection 18319
Junxi Chen (University of Chinese Academy of Sciences), Liang Li (Institute of computing technology, CAS), Li Su (University of Chinese Academy of Sciences), Zheng-jun Zha (University of Science and Technology of China), and Qingming Huang (University of Chinese Academy of Sciences)

Context-Guided Spatio-Temporal Video Grounding 18330
Xin Gu (University of Chinese Academy of Sciences, China), Heng Fan (University of North Texas, USA), Yan Huang (University of North Texas, USA), Tiejian Luo (University of Chinese Academy of Sciences, China), and Libo Zhang (Chinese Academy of Sciences, China)

Just Add π! Pose Induced Video Transformers for Understanding Activities of Daily Living 18340
Dominick Reilly (University of North Carolina at Charlotte) and Srijan Das (University of North Carolina at Charlotte)

Action Detection via an Image Diffusion Process 18351
Lin Geng Foo (Singapore University of Technology and Design (SUTD), Singapore), Tianjiao Li (Singapore University of Technology and Design (SUTD), Singapore), Hossein Rahmani (Lancaster University, United Kingdom), and Jun Liu (Singapore University of Technology and Design (SUTD), Singapore)

LLMs are Good Sign Language Translators 18362
Jia Gong (Singapore University of Technology and Design (SUTD), Singapore), Lin Geng Foo (Singapore University of Technology and Design (SUTD), Singapore), Yixuan He (Singapore University of Technology and Design (SUTD), Singapore), Hossein Rahmani (Lancaster University, United Kingdom), and Jun Liu (Singapore University of Technology and Design (SUTD), Singapore)

End-to-End Spatio-Temporal Action Localisation with Video Transformers 18373
Alexey A. Gritsenko (Google), Xuehan Xiong (Google), Josip Djolonga (Google), Mostafa Dehghani (Google), Chen Sun (Google), Mario Lucic (Google), Cordelia Schmid (Google), and Anurag Arnab (Google)

HIG: Hierarchical Interlacement Graph Approach to Scene Graph Generation in Video Understanding 18384
Trong-Thuan Nguyen (University of Arkansas, USA), Pha Nguyen (University of Arkansas, USA), and Khoa Luu (University of Arkansas, USA)

LLMs are Good Action Recognizers 18395
Haoxuan Qu (Singapore University of Technology and Design, Singapore), Yujun Cai (Nanyang Technological University, Singapore), and Jun Liu (Singapore University of Technology and Design, Singapore)

VideoLLM-online: Online Video Large Language Model for Streaming Video 18407
Joya Chen (National University of Singapore), Zhaoyang Lv (Reality Labs Research, Meta), Shiwei Wu (National University of Singapore), Kevin Qinghong Lin (National University of Singapore), Chenan Song (National University of Singapore), Difei Gao (National University of Singapore), Jia-Wei Liu (National University of Singapore), Ziteng Gao (National University of Singapore), Dongxing Mao (National University of Singapore), and Mike Zheng Shou (National University of Singapore)

What, When, and Where? Self-Supervised Spatio-Temporal Grounding in Untrimmed Multi-Action Videos from Narrated Instructions 18419
Brian Chen (Columbia University), Nina Shvetsova (Goethe University Frankfurt), Andrew Rouditchenko (MIT CSAIL), Daniel Kondermann (Quality Match GmbH), Samuel Thomas (IBM Research AI), Shih-Fu Chang (Columbia University), Rogerio Feris (IBM Research AI), James Glass (MIT CSAIL), and Hilde Kuehne (University of Bonn)

Narrative Action Evaluation with Prompt-Guided Multimodal Interaction 18430
Shiyi Zhang (Tsinghua University, China), Sule Bai (Tsinghua University, China), Guangyi Chen (Carnegie Mellon University, USA), Lei Chen (Tsinghua University, China), Jiwen Lu (Tsinghua University, China), Junle Wang (Tencent), and Yansong Tang (Tsinghua University, China)

Realigning Confidence with Temporal Saliency Information for Point-Level Weakly-Supervised Temporal Action Localization 18440
Ziying Xia (University of Electronic Science and Technology of China, China), Jian Cheng (University of Electronic Science and Technology of China, China), Siyu Liu (University of Electronic Science and Technology of China, China), Yongxiang Hu (University of Electronic Science and Technology of China, China), Shiguang Wang (University of Electronic Science and Technology of China, China), Yijie Zhang (University of Electronic Science and Technology of China, China), and Liwan Dang (University of Electronic Science and Technology of China, China; The Second Research Institute of Civil Aviation Administration of China, China)

Action-slot: Visual Action-centric Representations for Multi-label Atomic Activity Recognition in Traffic Scenes 18451
Chi-Hsi Kung (National Yang Ming Chiao Tung University, Taiwan), Shu-Wei Lu (National Yang Ming Chiao Tung University, Taiwan), Yi-Hsuan Tsai (Google, USA), and Yi-Ting Chen (National Yang Ming Chiao Tung University, Taiwan)

LoCoNet: Long-Short Context Network for Active Speaker Detection 18462
Xizi Wang (Indiana University Bloomington), Feng Cheng (UNC Chapel Hill), and Gedas Bertasius (UNC Chapel Hill)

Neighbor Relations Matter in Video Scene Detection 18473
Jiawei Tan (Chongqing University, China), Hongxing Wang (Chongqing University, China), Jiaxin Li (Chongqing University, China), Zhilong Ou (Chongqing University, China), and Zhangbin Qian (Chongqing University, China)

PREGO: Online Mistake Detection in PRocedural EGOcentric Videos 18483
Alessandro Flaborea (Sapienza University of Rome, Italy), Guido Maria D'Amely di Melendugno (Sapienza University of Rome, Italy), Leonardo Plini (Sapienza University of Rome, Italy), Luca Scofano (Sapienza University of Rome, Italy), Edoardo De Matteis (Sapienza University of Rome, Italy), Antonino Furnari (University of Catania, Italy), Giovanni Maria Farinella (University of Catania, Italy), and Fabio Galasso (Sapienza University of Rome, Italy)

Learning Object State Changes in Videos: An Open-World Perspective 18493
Zihui Xue (UT Austin, Meta), Kumar Ashutosh (UT Austin, Meta), and Kristen Grauman (UT Austin, Meta)

Enhanced Motion-Text Alignment for Image-to-Video Transfer Learning 18504
Wei Zhang (University of Science and Technology of China, China), Chaoqun Wan (Alibaba Cloud, China), Tongliang Liu (The University of Sydney, Australia), Xinmei Tian (University of Science and Technology of China, China), Xu Shen (Alibaba Cloud, China), and Jieping Ye (Alibaba Cloud, China)

Asymmetric Masked Distillation for Pre-Training Small Foundation Models 18516
Zhiyu Zhao (Nanjing University, China; Shanghai AI Lab, China), Bingkun Huang (Nanjing University, China; Shanghai AI Lab, China), Sen Xing (Shanghai AI Lab, China), Gangshan Wu (Nanjing University, China), Yu Qiao (Shanghai AI Lab, China), and Limin Wang (Nanjing University, China; Shanghai AI Lab, China)

Harnessing Large Language Models for Training-free Video Anomaly Detection 18527
Luca Zanella (University of Trento, Italy), Willi Menapace (University of Trento, Italy), Massimiliano Mancini (University of Trento, Italy), Yiming Wang (Fondazione Bruno Kessler, Italy), and Elisa Ricci (University of Trento, Italy, Fondazione Bruno Kessler, Italy)

SportsHHI: A Dataset for Human-Human Interaction Detection in Sports Videos 18537
Tao Wu (Nanjing University, China), Runyu He (Nanjing University, China), Gangshan Wu (Nanjing University, China), and Limin Wang (Nanjing University, China; Shanghai AI lab, China)

VicTR: Video-conditioned Text Representations for Activity Recognition 18547
Kumara Kahatapitiya (Stony Brook University), Anurag Arnab (Google Research), Arsha Nagrani (Google Research), and Michael S. Ryoo (Stony Brook University, Google Research)

Dual DETRs for Multi-Label Temporal Action Detection 18559
Yuhan Zhu (Nanjing University, China), Guozhen Zhang (Nanjing University, China), Jing Tan (The Chinese University of Hong Kong, Hong Kong), Gangshan Wu (Nanjing University, China), and Limin Wang (Nanjing University, China; Shanghai AI Lab, China)

Adapting Short-Term Transformers for Action Detection in Untrimmed Videos 18570
Min Yang (Nanjing University, China), Huan Gao (Inchitech, China), Ping Guo (Intel Labs, China), and Limin Wang (Nanjing University, China; Shanghai AI Lab, China)

Can't Make an Omelette Without Breaking Some Eggs: Plausible Action Anticipation Using Large Video-Language Models 18580
Himangi Mittal (Carnegie Mellon University, USA), Nakul Agarwal (Honda Research Institute, USA), Shao-Yuan Lo (Honda Research Institute, USA), and Kwonjoon Lee (Honda Research Institute, USA)

End-to-End Temporal Action Detection with 1B Parameters Across 1000 Frames 18591
Shuming Liu (King Abdullah University of Science and Technology, Saudi Arabia), Chen-Lin Zhang (4Paradigm Inc, China), Chen Zhao (King Abdullah University of Science and Technology, Saudi Arabia), and Bernard Ghanem (King Abdullah University of Science and Technology, Saudi Arabia)

RMem: Restricted Memory Banks Improve Video Object Segmentation 18602
Junbao Zhou (University of Illinois Urbana-Champaign), Ziqi Pang (University of Illinois Urbana-Champaign), and Yu-Xiong Wang (University of Illinois Urbana-Champaign)

Low-power, Continuous Remote Behavioral Localization with Event Cameras 18612
Friedhelm Hamann (Technical University Berlin, Germany), Suman Ghosh (Technical University Berlin, Germany), Ignacio Juarez Martinez (Oxford University, UK), Tom Hart (Oxford Brookes University, UK), Alex Kacelnik (Oxford University, UK), and Guillermo Gallego (Technical University Berlin, Germany)

Action Scene Graphs for Long-Form Understanding of Egocentric Videos 18622
Ivan Rodin (University of Catania), Antonino Furnari (University of Catania), Kyle Min (Intel Labs), Subarna Tripathi (Intel Labs), and Giovanni Maria Farinella (University of Catania)

ExACT: Language-guided Conceptual Reasoning and Uncertainty Estimation for Event-based Action Recognition and More 18633
Jiazhou Zhou (The Hong Kong University of Science and Technology (Guangzhou), China), Xu Zheng (The Hong Kong University of Science and Technology (Guangzhou), China), Yuanhuiyi Lyu (The Hong Kong University of Science and Technology (Guangzhou), China), and Lin Wang (The Hong Kong University of Science and Technology (Guangzhou), China, The Hong Kong University of Science and Technology, China)

Uncertainty-aware Action Decoupling Transformer for Action Anticipation 18644
Hongji Guo (Rensselaer Polytechnic Institute), Nakul Agarwal (Honda Research Institute, USA), Shao-Yuan Lo (Honda Research Institute, USA), Kwonjoon Lee (Honda Research Institute, USA), and Qiang Ji (Rensselaer Polytechnic Institute)

Error Detection in Egocentric Procedural Task Videos 18655
Shih-Po Lee (Northeastern University, USA), Zijia Lu (Northeastern University, USA), Zekun Zhang (Stony Brook University, USA), Minh Hoai (Stony Brook University, USA), and Ehsan Elhamifar (Northeastern University, USA)

Learning to Predict Activity Progress by Self-Supervised Video Alignment 18667
Gerard Donahue (Northeastern University, USA) and Ehsan Elhamifar (Northeastern University, USA)

MaskCLR: Attention-Guided Contrastive Learning for Robust Action Representation Learning . 18678
Mohamed Abdelfattah (EPFL), Mariam Hassan (EPFL), and Alexandre Alahi (EPFL)

Align Before Adapt: Leveraging Entity-to-Region Alignments for Generalizable Video Action Recognition 18688
Yifei Chen (Huawei Technologies), Dapeng Chen (Huawei Technologies), Ruijin Liu (Huawei Technologies), Sai Zhou (Huawei Technologies), Wenyuan Xue (Huawei Technologies), and Wei Peng (Huawei Technologies)

DIBS: Enhancing Dense Video Captioning with Unlabeled Videos via Pseudo Boundary Enrichment and Online Refinement 18699
Hao Wu (University of Science and Technology of China;Shanghai Artificial Intelligence Laboratory), Huabin Liu (Shanghai Artificial Intelligence Laboratory;Shanghai Jiao Tong University), Yu Qiao (Shanghai Artificial Intelligence Laboratory), and Xiao Sun (Shanghai Artificial Intelligence Laboratory)

Bridging the Gap: A Unified Video Comprehension Framework for Moment Retrieval and Highlight Detection 18709
Yicheng Xiao (Tsinghua University, China), Zhuoyan Luo (Tsinghua University, China), Yong Liu (Tsinghua University, China), Yue Ma (Tsinghua University, China), Hengwei Bian (Carnegie Mellon University, America), Yatai Ji (Tsinghua University, China), Yujiu Yang (Tsinghua University, China), and Xiu Li (Tsinghua University, China)

Test-Time Zero-Shot Temporal Action Localization 18720
Benedetta Liberatori (University of Trento, Italy), Alessandro Conti (University of Trento, Italy), Paolo Rota (University of Trento, Italy), Yiming Wang (Fondazione Bruno Kessler, Italy), and Elisa Ricci (University of Trento, Italy, Fondazione Bruno Kessler, Italy)

Selective, Interpretable and Motion Consistent Privacy Attribute Obfuscation for Action Recognition 18730
Filip Ilic (Graz University of Technology, Austria), He Zhao (York University, Canada), Thomas Pock (Graz University of Technology, Austria), and Richard P. Wildes (York University, Canada)

Step Differences in Instructional Video 18740
Tushar Nagarajan (FAIR, Meta) and Lorenzo Torresani (FAIR, Meta)

Compositional Video Understanding with Spatiotemporal Structure-based Transformers 18751
Hoyeoung Yun (Hanyang University), Jinwoo Ahn (Hanyang University), Minseo Kim (Hanyang University), and Eun-Sol Kim (Hanyang University)

Part-aware Unified Representation of Language and Skeleton for Zero-shot Action Recognition 18761
Anqi Zhu (The University of Melbourne), Qiuhong Ke (Monash University), Mingming Gong (The University of Melbourne), and James Bailey (The University of Melbourne)

vid-TLDR: Training Free Token Merging for Light-weight Video Transformer 18771
Joonmyung Choi (Korea University, South Korea), Sanghyeok Lee (Korea University, South Korea), Jaewon Chu (Korea University, South Korea), Minhyuk Choi (Korea University, South Korea), and Hyunwoo J. Kim (Korea University, South Korea)

CPR-Coach: Recognizing Composite Error Actions based on Single-class Training 18782
Shunli Wang (Fudan University, China), Shuaibing Wang (Fudan University, China), Dingkang Yang (Fudan University, China), Mingcheng Li (Fudan University, China), Haopeng Kuang (Fudan University, China), Xiao Zhao (Fudan University, China), Liuzhen Su (Fudan University, China), Peng Zhai (Fudan University, China), and Lihua Zhang (Fudan University, China)

Uncovering What, Why and How: A Comprehensive Benchmark for Causation Understanding of Video Anomaly 18793
Hang Du (Beijing University of Posts and Telecommunications, China), Sicheng Zhang (Beijing University of Posts and Telecommunications, China), Binzhu Xie (Beijing University of Posts and Telecommunications, China), Guoshun Nan (Beijing University of Posts and Telecommunications, China), Jiayang Zhang (Beijing University of Posts and Telecommunications, China), Junrui Xu (Beijing University of Posts and Telecommunications, China), Hangyu Liu (Beijing University of Posts and Telecommunications, China), Sicong Leng (Nanyang Technological University, Singapore), Jiangming Liu (Yunnan University, China), Hehe Fan (Zhejiang University, China), Dajiu Huang (China Telecom Co., Ltd. Sichuan Branch, China), Jing Feng (China Telecom Co., Ltd. Sichuan Branch, China), Linli Chen (China Telecom Co., Ltd. Sichuan Branch, China), Can Zhang (Beijing University of Posts and Telecommunications, China), Xuhuan Li (Beijing University of Posts and Telecommunications, China), Hao Zhang (Beijing University of Posts and Telecommunications, China), Jianhang Chen (Beijing University of Posts and Telecommunications, China), Qimei Cui (Beijing University of Posts and Telecommunications, China), and Xiaofeng Tao (Beijing University of Posts and Telecommunications, China)

Detours for Navigating Instructional Videos 18804
Kumar Ashutosh (UT Austin and FAIR, Meta), Zihui Xue (UT Austin and FAIR, Meta), Tushar Nagarajan (FAIR, Meta), and Kristen Grauman (UT Austin and FAIR, Meta)

Why Not Use Your Textbook? Knowledge-Enhanced Procedure Planning of Instructional Videos 18816
Kumaranage Ravindu Yasas Nagasinghe (Mohamed bin Zayed University of Artificial Intelligence), Honglu Zhou (NEC Laboratories, USA), Malitha Gunawardhana (University of Auckland), Martin Renqiang Min (NEC Laboratories, USA), Daniel Harari (Weizmann Institute of Science), and Muhammad Haris Khan (Mohamed bin Zayed University of Artificial Intelligence)

Multiscale Vision Transformers Meet Bipartite Matching for Efficient Single-stage Action Localization 18827
Ioanna Ntinou (Queen Mary University of London), Enrique Sanchez (Samsung AI Center Cambridge), and Georgios Tzimiropoulos (Queen Mary University London; Samsung AI Center Cambridge)

TE-TAD: Towards Full End-to-End Temporal Action Detection via Time-Aligned Coordinate Expression 18837
Ho-Joong Kim (Korea University), Jung-Ho Hong (Korea University), Heejo Kong (Korea University), and Seong-Whan Lee (Korea University)

CSTA: CNN-based Spatiotemporal Attention for Video Summarization 18847
Jaewon Son (Sungkyunkwan University, Republic of Korea), Jaehun Park (Sungkyunkwan University, Republic of Korea), and Kwangsu Kim (Sungkyunkwan University, Republic of Korea)

PeVL: Pose-Enhanced Vision-Language Model for Fine-Grained Human Action Recognition 18857
*Haosong Zhang (Nanyang Technological University, Singapore), Mei Chee Leong (Institute for Infocomm Research (I2R), A*STAR, Singapore), Liyuan Li (Institute for Infocomm Research (I2R), A*STAR, Singapore), and Weisi Lin (Nanyang Technological University, Singapore)*

MULDE: Multiscale Log-Density Estimation via Denoising Score Matching for Video Anomaly Detection 18868
Jakub Micorek (Graz University of Technology, Austria), Horst Possegger (Graz University of Technology, Austria), Dominik Narnhofer (Graz University of Technology, Austria), Horst Bischof (Graz University of Technology, Austria), and Mateusz Kozinski (Graz University of Technology, Austria)

Language Model Guided Interpretable Video Action Reasoning 18878
Ning Wang (Xidian University), Guangming Zhu (Xidian University), HS Li (Xidian University), Liang Zhang (Xidian University), Syed Afaq Ali Shah (Edith Cowan University), and Mohammed Bennamoun (University of Western Australia)

OST: Refining Text Knowledge with Optimal Spatio-Temporal Descriptor for General Video Recognition 18888
Tongjia Chen (Hunan University, China), Hongshan Yu (Hunan University, China), Zhengeng Yang (Hunan Normal University, China), Zechuan Li (Hunan University, China), Wei Sun (Hunan University, China), and Chen Chen (University of Central Florida, USA)

Text Prompt with Normality Guidance for Weakly Supervised Video Anomaly Detection 18899
Zhiwei Yang (Guangzhou Institute of Technology, Xidian University, Guangzhou, China), Jing Liu (Guangzhou Institute of Technology, Xidian University, Guangzhou, China), and Peng Wu (School of Computer Science, Northwestern Polytechnical University, Xi'an, China)

VideoGrounding-DINO: Towards Open-Vocabulary Spatio-Temporal Video Grounding 18909
Syed Talal Wasim (Mohamed bin Zayed University of AI, UAE), Muzammal Naseer (Mohamed bin Zayed University of AI, UAE), Salman Khan (Mohamed bin Zayed University of AI, UAE; Australian National University, Australia), Ming-Hsuan Yang (University of California, Merced, USA; Google Research, USA), and Fahad Shahbaz Khan (Mohamed bin Zayed University of AI, UAE; Linkoping University, Sweden)

Unsupervised Video Domain Adaptation with Masked Pre-Training and Collaborative Self-Training 18919
Arun Reddy (Johns Hopkins University Applied Physics Laboratory), William Paul (Johns Hopkins University Applied Physics Laboratory), Corban Rivera (Johns Hopkins University Applied Physics Laboratory), Ketul Shah (Johns Hopkins University), Celso M. de Melo (Army Research Laboratory), and Rama Chellappa (Johns Hopkins University)

SnAG: Scalable and Accurate Video Grounding 18930
Fangzhou Mu (University of Wisconsin - Madison), Sicheng Mo (University of Wisconsin - Madison), and Yin Li (University of Wisconsin - Madison)

Learning Correlation Structures for Vision Transformers 18941
Manjin Kim (Pohang University of Science and Technology (POSTECH), Republic of Korea), Paul Hongsuck Seo (Korea University, Republic of Korea), Cordelia Schmid (Google Research, France), and Minsu Cho (Pohang University of Science and Technology (POSTECH), Republic of Korea)

Weakly-Supervised Audio-Visual Video Parsing with Prototype-based Pseudo-Labeling 18952
Kranthi Kumar Rachavarapu (Indian Institute of Technology Madras), Kalyan Ramakrishnan (University of Oxford), and Rajagopalan A. N. (Indian Institute of Technology Madras)

Matching Anything by Segmenting Anything 18963
Siyuan Li (ETH Zurich), Lei Ke (ETH Zurich), Martin Danelljan (ETH Zurich), Luigi Piccinelli (ETH Zurich), Mattia Segu (ETH Zurich), Luc Van Gool (ETH Zurich, INSAIT), and Fisher Yu (ETH Zurich)

3D Feature Tracking via Event Camera 18974
Siqi Li (Tsinghua University, China), Zhikuan Zhou (Tsinghua University, China), Zhou Xue (Li Auto, China), Yipeng Li (Tsinghua University, China), Shaoyi Du (Xi'an Jiaotong University, China), and Yue Gao (Tsinghua University, China)

Frequency Decoupling for Motion Magnification via Multi-Level Isomorphic Architecture 18984
Fei Wang (Hefei University of Technology), Dan Guo (Hefei University of Technology; Hefei Comprehensive National Science Center), Kun Li (Hefei University of Technology), Zhun Zhong (Hefei University of Technology; University of Nottingham), and Meng Wang (Hefei University of Technology; Hefei Comprehensive National Science Center)

Towards Generalizable Multi-Object Tracking 18995
Zheng Qin (Xi'an Jiaotong University, China), Le Wang (Xi'an Jiaotong University, China), Sanping Zhou (Xi'an Jiaotong University, China), Panpan Fu (Xi'an Jiaotong University, China), Gang Hua (Wormpex AI Research, China), and Wei Tang (University of Illinois, USA)

SocialCircle: Learning the Angle-based Social Interaction Representation for Pedestrian Trajectory Prediction 19005
Conghao Wong (Huazhong University of Science and Technology), Beihao Xia (Huazhong University of Science and Technology), Ziqian Zou (Huazhong University of Science and Technology), Yulong Wang (Huazhong Agricultural University), and Xinge You (Huazhong University of Science and Technology)

Self-Supervised Multi-Object Tracking with Path Consistency 19016
Zijia Lu (Northeastern University, USA), Bing Shuai (Amazon, USA), Yanbei Chen (Amazon, USA), Zhenlin Xu (Amazon, USA), and Davide Modolo (Amazon, USA)

UnSAMFlow: Unsupervised Optical Flow Guided by Segment Anything Model 19027
Shuai Yuan (Meta Reality Labs), Lei Luo (Meta Reality Labs), Zhuo Hui (Meta Reality Labs), Can Pu (Meta Reality Labs), Xiaoyu Xiang (Meta Reality Labs), Rakesh Ranjan (Meta Reality Labs), and Denis Demandolx (Meta Reality Labs)

RTracker: Recoverable Tracking via PN Tree Structured Memory 19038
Yuqing Huang (Harbin Institute of Technology, Shenzhen; Peng Cheng Laboratory), Xin Li (Peng Cheng Laboratory), Zikun Zhou (Peng Cheng Laboratory), Yaowei Wang (Peng Cheng Laboratory), Zhenyu He (Harbin Institute of Technology, Shenzhen), and Ming-Hsuan Yang (UC Merced; Yonsei University)

ARTrackV2: Prompting Autoregressive Tracker Where to Look and How to Describe 19048
Yifan Bai (Xi'an Jiaotong University, China), Zeyang Zhao (Xi'an Jiaotong University, China), Yihong Gong (Xi'an Jiaotong University, China), and Xing Wei (Xi'an Jiaotong University, China)

Endow SAM with Keen Eyes: Temporal-spatial Prompt Learning for Video Camouflaged Object Detection 19058
Wenjun Hui (Beijing Jiaotong University, China), Zhenfeng Zhu (Beijing Jiaotong University, China), Shuai Zheng (Beijing Jiaotong University, China), and Yao Zhao (Beijing Jiaotong University, China)

MemFlow: Optical Flow Estimation and Prediction with Memory 19068
Qiaole Dong (Fudan University, China) and Yanwei Fu (Fudan University, China)

OneTracker: Unifying Visual Object Tracking with Foundation Models and Efficient Tuning 19079
Lingyi Hong (Fudan University, China), Shilin Yan (Fudan University, China), Renrui Zhang (The Chinese University of Hong Kong), Wanyun Li (Fudan University, China), Xinyu Zhou (Fudan University, China), Pinxue Guo (Fudan University, China), Kaixun Jiang (Fudan University, China), Yiting Chen (Fudan University, China), Jinglun Li (Fudan University, China), Zhaoyu Chen (Fudan University, China), and Wenqiang Zhang (Fudan University, China)

Learned Trajectory Embedding for Subspace Clustering 19092
Yaroslava Lochman (Chalmers University of Technology), Carl Olsson (Chalmers University of Technology, Lund University), and Christopher Zach (Chalmers University of Technology)

PNeRV: Enhancing Spatial Consistency via Pyramidal Neural Representation for Videos 19103
Qi Zhao (Nanjing University, China), M. Salman Asif (University of California Riverside, USA), and Zhan Ma (Nanjing University, China)

DiffusionTrack: Point Set Diffusion Model for Visual Object Tracking 19113
Fei Xie (Shanghai Jiao Tong University), Zhongdao Wang (Huawei Noah's Ark Lab), and Chao Ma (Shanghai Jiao Tong University)

Sparse Global Matching for Video Frame Interpolation with Large Motion 19125
Chunxu Liu (Nanjing University, China), Guozhen Zhang (Nanjing University, China), Rui Zhao (SenseTime Research, China), and Limin Wang (Nanjing University, China; Shanghai AI Lab, China)

iKUN: Speak to Trackers without Retraining 19135
Yunhao Du (Beijing University of Posts and Telecommunications, China), Cheng Lei (Beijing University of Posts and Telecommunications, China), Zhicheng Zhao (Beijing University of Posts and Telecommunications, China), and Fei Su (Beijing University of Posts and Telecommunications, China)

NetTrack: Tracking Highly Dynamic Objects with a Net 19145
Guangze Zheng (The University of Hong Kong, Centre for Transformative Garment Production), Shijie Lin (The University of Hong Kong, Centre for Transformative Garment Production), Haobo Zuo (The University of Hong Kong, Centre for Transformative Garment Production), Changhong Fu (Tongji University), and Jia Pan (The University of Hong Kong, Centre for Transformative Garment Production)

Single-Model and Any-Modality for Video Object Tracking 19156
Zongwei Wu (CVL, University of Wurzburg), Jilai Zheng (AI Institute, Shanghai Jiao Tong University), Xiangxuan Ren (AI Institute, Shanghai Jiao Tong University), Florin-Alexandru Vasluianu (CVL, University of Wurzburg), Chao Ma (AI Institute, Shanghai Jiao Tong University), Danda Pani Paudel (INSAIT Sophia), Luc Van Gool (CVL, ETHz; INSAIT Sophia), and Radu Timofte (CVL, University of Wurzburg)

FlowDiffuser: Advancing Optical Flow Estimation with Diffusion Models 19167
Ao Luo (Southwest Jiaotong University, China), Xin Li (G42, United Arab Emirates), Fan Yang (G42, United Arab Emirates), Jiangyu Liu (Megvii Technology, China), Haoqiang Fan (Megvii Technology, China), and Shuaicheng Liu (University of Electronic Science and Technology of China, China)

Video Harmonization with Triplet Spatio-Temporal Variation Patterns 19177
Zonghui Guo (Institute of Computing Technology, Chinese Academy of Sciences), Xinyu Han (Ocean University of China), Jie Zhang (Institute of Computing Technology, Chinese Academy of Sciences), Shiguang Shan (Institute of Computing Technology, Chinese Academy of Sciences), and Haiyong Zheng (Ocean University of China)

Dense Optical Tracking: Connecting the Dots 19187
Guillaume Le Moing (Inria, Paris, France), Jean Ponce (Ecole normale supérieure, Paris, France), and Cordelia Schmid (Inria, Paris, France)

Efficient Meshflow and Optical Flow Estimation from Event Cameras 19198
Xinglong Luo (University of Electronic Science and Technology of China; Megvii Technology), Ao Luo (Southwest Jiaotong University; Megvii Technology), Zhengning Wang (University of Electronic Science and Technology of China), Chunyu Lin (Beijing Jiaotong University), Bing Zeng (University of Electronic Science and Technology of China), and Shuaicheng Liu (University of Electronic Science and Technology of China; Megvii Technology)

Context-Aware Integration of Language and Visual References for Natural Language Tracking . 19208
Yanyan Shao (Zhejiang University of Technology, China), Shuting He (Nanyang Technological University, Singapore), Qi Ye (Zhejiang University, China), Yuchao Feng (Zhejiang University of Technology, China), Wenhan Luo (The Hong Kong University of Science and Technology, China), and Jiming Chen (Zhejiang University, China; Zhejiang University of Technology, China)

Depth-aware Test-Time Training for Zero-shot Video Object Segmentation 19218
Weihuang Liu (University of Macau), Xi Shen (Intellindust), Haolun Li (University of Macau), Xiuli Bi (Chongqing University of Posts and Telecommunications), Bo Liu (Chongqing University of Posts and Telecommunications), Chi-Man Pun (University of Macau), and Xiaodong Cun (Tencent AI Lab)

Weakly Supervised Video Individual Counting 19228
Xinyan Liu (University of Chinese Academy of Science), Guorong Li (University of Chinese Academy of Science; Key Lab of Big Data Mining and Knowledge Management, UCAS, China; Key Lab of Intell. Info. Process., Inst. of Comput. Tech., CAS;), Yuankai Qi (Macquarie University), Ziheng Yan (University of Chinese Academy of Science), Zhenjun Han (University of Chinese Academy of Science;), Anton van den Hengel (Australian Institute for Machine Learning, The University of Adelaide), Ming-Hsuan Yang (University of California at Merced), and Qingming Huang (University of Chinese Academy of Science; Key Lab of Big Data Mining and Knowledge Management, UCAS, China; Key Lab of Intell. Info. Process., Inst. of Comput. Tech., CAS;)

Dual Prototype Attention for Unsupervised Video Object Segmentation 19238
Suhwan Cho (Yonsei University), Minhyeok Lee (Yonsei University), Seunghoon Lee (Yonsei University), Dogyoon Lee (Yonsei University), Heeseung Choi (Korea Institute of Science and Technology (KIST)), Ig-Jae Kim (Korea Institute of Science and Technology (KIST)), and Sangyoun Lee (Yonsei University)

Event Stream-based Visual Object Tracking: A High-Resolution Benchmark Dataset and A Novel Baseline 19248
Xiao Wang (Anhui University, China), Shiao Wang (Anhui University, China), Chuanming Tang (University of Chinese Academy of Sciences, China), Lin Zhu (Beijing Institute of Technology, China), Bo Jiang (Anhui University, China), Yonghong Tian (Peking University, China), and Jin Tang (Anhui University, China)

HIPTrack: Visual Tracking with Historical Prompts 19258
Wenrui Cai (Beihang University, China), Qingjie Liu (Beihang University, China), and Yunhong Wang (Beihang University, China)

FlowTrack: Revisiting Optical Flow for Long-Range Dense Tracking 19268
Seokju Cho (Korea University), Jiahui Huang (Adobe Research), Seungryong Kim (Korea University), and Joon-Young Lee (Adobe Research)

Implicit Motion Function 19278
Yue Gao (Microsoft Research, China), Jiahao Li (Microsoft Research, China), Lei Chu (Microsoft Research, China), and Yan Lu (Microsoft Research, China)

DeconfuseTrack: Dealing with Confusion for Multi-Object Tracking 19290
Cheng Huang (Huazhong University of Science and Technology), Shoudong Han (Huazhong University of Science and Technology), Mengyu He (Huazhong University of Science and Technology), Wenbo Zheng (Huazhong University of Science and Technology), and Yuhao Wei (Huazhong University of Science and Technology)

Autoregressive Queries for Adaptive Tracking with Spatio-Temporal Transformers 19300
Jinxia Xie (Guangxi Normal University, China), Bineng Zhong (Guangxi Normal University, China), Zhiyi Mo (Wuzhou University, China), Shengping Zhang (Harbin Institute of Technology), Liangtao Shi (Guangxi Normal University, China), Shuxiang Song (Guangxi Normal University, China), and Rongrong Ji (Xiamen University, China)

ExtDM: Distribution Extrapolation Diffusion Model for Video Prediction 19310
Zhicheng Zhang (VCIP & TMCC & DISSec, Nankai University,Nankai International Advanced Research Institute (SHENZHEN· FUTIAN)), Junyao Hu (VCIP & TMCC & DISSec, Nankai University,Nankai International Advanced Research Institute (SHENZHEN· FUTIAN)), Wentao Cheng (VCIP & TMCC & DISSec, Nankai University), Danda Paudel (Computer Vision Lab, ETH Zurich,INSAIT, Sofia University), and Jufeng Yang (VCIP & TMCC & DISSec, Nankai University,Nankai International Advanced Research Institute (SHENZHEN· FUTIAN))

DiffMOT: A Real-time Diffusion-based Multiple Object Tracker with Non-linear Prediction 19321
Weiyi Lv (Shanghai University, China), Yuhang Huang (National University of Defense Technology, China), Ning Zhang (PAII Inc., USA), Ruei-Sung Lin (PAII Inc., USA), Mei Han (PAII Inc., USA), and Dan Zeng (Shanghai University, China)

GigaTraj: Predicting Long-term Trajectories of Hundreds of Pedestrians in Gigapixel Complex Scenes 19331
Haozhe Lin (Tsinghua University), Chunyu Wei (Tsinghua University), Li He (Tsinghua University), Yuchen Guo (Tsinghua University), Yunqi Zhao (Tsinghua University), Shanglong Li (Tsinghua University), and Lu Fang (Tsinghua University)

Delving into the Trajectory Long-tail Distribution for Muti-object Tracking 19341
Sijia Chen (Huazhong University of Science and Technology, China), En Yu (Huazhong University of Science and Technology, China), Jinyang Li (Huazhong University of Science and Technology, China), and Wenbing Tao (Huazhong University of Science and Technology, China)

OCAI: Improving Optical Flow Estimation by Occlusion and Consistency Aware Interpolation .. 19352
Jisoo Jeong (Qualcomm AI Research), Hong Cai (Qualcomm AI Research), Risheek Garrepalli (Qualcomm AI Research), Jamie Menjay Lin (Qualcomm AI Reseach), Munawar Hayat (Qualcomm AI Research), and Fatih Porikli (Qualcomm AI Research)

Orals 5A Datasets and Evaluation

Deep Generative Model based Rate-Distortion for Image Downscaling Assessment 19363
Yuanbang Liang (Cardiff University, UK), Bhavesh Garg (IIT Bombay & WadhwaniAI, India), Paul Rosin (Cardiff University, UK), and Yipeng Qin (Cardiff University, UK)

360+x: A Panoptic Multi-modal Scene Understanding Dataset 19373
Hao Chen (University of Birmingham, UK), Yuqi Hou (University of Birmingham, UK), Chenyuan Qu (University of Birmingham, UK), Irene Testini (University of Birmingham, UK), Xiaohan Hong (University of Birmingham, UK), and Jianbo Jiao (University of Birmingham, UK)

Ego-Exo4D: Understanding Skilled Human Activity from First- and Third-Person Perspectives . 19383
Kristen Grauman (FAIR, Meta; University of Texas at Austin, USA), Andrew Westbury (FAIR, Meta, USA), Lorenzo Torresani (FAIR, Meta, USA), Kris Kitani (FAIR, Meta and Carnegie Mellon University, USA), Jitendra Malik (FAIR, Meta; University of California, Berkeley, USA), Triantafyllos Afouras (FAIR, Meta, USA), Kumar Ashutosh (FAIR, Meta; University of Texas at Austin, USA), Vijay Baiyya (Meta, USA), Siddhant Bansal (University of Bristol, UK), Bikram Boote (University of Illinois, Urbana Champaign, USA), Eugene Byrne (FAIR, Meta; Carnegie Mellon University, USA), Zach Chavis (University of Minnesota, USA), Joya Chen (National University of Singapore, Singapore), Feng Cheng (FAIR, Meta, USA), Fu-Jen Chu (FAIR, Meta, USA), Sean Crane (Carnegie Mellon University, USA), Avijit Dasgupta (International Institute of Information Technology, Hyderabad, India), Jing Dong (Meta, USA), Maria Escobar (Universidad de los Andes, Colombia), Cristhian Forigua (Universidad de los Andes, Colombia), Abrham Gebreselasie (Carnegie Mellon University, Rwanda), Sanjay Haresh (Simon Fraser University, Canada), Jing Huang (FAIR, Meta, USA), Md Mohaiminul Islam (University of North Carolina, Chapel Hill, USA), Suyog Jain (FAIR, Meta, USA), Rawal Khirodkar (Carnegie Mellon University, USA), Devansh Kukreja (FAIR, Meta, USA), Kevin J Liang (FAIR, Meta, USA), Jia-Wei Liu (National University of Singapore, Singapore), Sagnik Majumder (FAIR, Meta; University of Texas at Austin, USA), Yongsen Mao (Simon Fraser University, Canada), Miguel Martin (FAIR, Meta, USA), Effrosyni Mavroudi (FAIR, Meta, USA), Tushar Nagarajan (FAIR, Meta, USA), Francesco Ragusa (University of Catania, Italy), Santhosh Kumar Ramakrishnan (University of Texas at Austin, USA), Luigi Seminara (University of Catania, Italy), Arjun Somayazulu (University of Texas at Austin, USA), Yale Song (FAIR, Meta, USA), Shan Su (University of Pennsylvania, USA), Zihui Xue (FAIR, Meta; University of Texas at Austin, USA), Edward Zhang (University of Pennsylvania, USA), Jinxu Zhang (University of Pennsylvania, USA), Angela Castillo (Universidad de los Andes, Colombia), Changan Chen (University of Texas at Austin, USA), Xinzhu Fu (National University of Singapore, Singapore), Ryosuke Furuta (University of Tokyo, Japan), Cristina Gonzalez (Universidad de los Andes, Colombia), Prince Gupta (Meta, USA), Jiabo Hu (Meta, USA), Yifei Huang (University of Tokyo, Japan), Yiming Huang (University of Pennsylvania, USA), Weslie Khoo (Indiana University, USA), Anush Kumar (University of Minnesota, USA), Robert Kuo (Meta, USA), Sach Lakhavani (Meta, USA), Miao Liu (Meta, USA), Mi Luo (University of Texas at Austin, USA), Zhengyi Luo (Carnegie Mellon University, USA), Brighid Meredith (Meta, USA), Austin Miller (Meta, USA), Oluwatumininu Oguntola (University of North Carolina, Chapel Hill, USA), Xiaqing Pan (Meta, USA), Penny Peng (Meta, USA), Shraman Pramanick (Johns Hopkins University, USA), Merey Ramazanova (KAUST, Saudi Arabia), Fiona Ryan (Georgia Tech, USA), Wei Shan (University of North Carolina, Chapel Hill, USA), Kiran Somasundaram (Meta, USA), Chenan Song (National University of Singapore, Singapore), Audrey Southerland (Georgia Tech, USA), Masatoshi Tateno (University of Tokyo, Japan), Huiyu Wang (FAIR, Meta, USA), Yuchen Wang (Indiana University, USA), Takuma Yagi (University of Tokyo, Japan), Mingfei Yan (Meta, USA), Xitong Yang (FAIR, Meta,

USA), Zecheng Yu (University of Tokyo, Japan), Shengxin Cindy Zha (Meta, USA), Chen Zhao (KAUST, Saudi Arabia), Ziwei Zhao (Indiana University, USA), Zhifan Zhu (University of Bristol, UK), Jeff Zhuo (University of North Carolina, Chapel Hill, USA), Pablo Arbelaez (Universidad de los Andes, Colombia), Gedas Bertasius (University of North Carolina, Chapel Hill, USA), Dima Damen (University of Bristol, UK), Jakob Engel (Meta, USA), Giovanni Maria Farinella (University of Catania, Italy), Antonino Furnari (University of Catania, Italy), Bernard Ghanem (KAUST, Saudi Arabia), Judy Hoffman (Georgia Tech, USA), C.V. Jawahar (International Institute of Information Technology, Hyderabad, India), Richard Newcombe (Meta, USA), Hyun Soo Park (University of Minnesota, USA), James M. Rehg (University of Illinois, Urbana Champaign, USA), Yoichi Sato (University of Tokyo, Japan), Manolis Savva (Simon Fraser University, Canada), Jianbo Shi (University of Pennsylvania, USA), Mike Zheng Shou (National University of Singapore, Singapore), and Michael Wray (University of Bristol, UK)

Rich Human Feedback for Text-to-Image Generation 19401
Youwei Liang (University of California, San Diego), Junfeng He (Google Research), Gang Li (Google Research), Peizhao Li (Brandeis University), Arseniy Klimovskiy (Google Research), Nicholas Carolan (Google Research), Jiao Sun (University of Southern California), Jordi Pont-Tuset (Google Research), Sarah Young (Google Research), Feng Yang (Google Research), Junjie Ke (Google Research), Krishnamurthy Dj Dvijotham (Google Research), Katherine M. Collins (University of Cambridge), Yiwen Luo (Google Research), Yang Li (Google Research), Kai J Kohlhoff (Google Research), Deepak Ramachandran (Google Research), and Vidhya Navalpakkam (Google Research)

BioCLIP: A Vision Foundation Model for the Tree of Life 19412
Samuel Stevens (The Ohio State University), Jiaman Wu (The Ohio State University), Matthew J Thompson (The Ohio State University), Elizabeth G Campolongo (The Ohio State University), Chan Hee Song (The Ohio State University), David Edward Carlyn (The Ohio State University), Li Dong (Microsoft Research), Wasila M Dahdul (University of California, Irvine), Charles Stewart (Rensselaer Polytechnic Institute), Tanya Berger-Wolf (The Ohio State University), Wei-Lun Chao (The Ohio State University), and Yu Su (The Ohio State University)

Orals 5B 3D from Multiview and Sensors

Grounding and Enhancing Grid-based Models for Neural Fields 19425
Zelin Zhao (Shanghai Jiao Tong University, China), Fenglei Fan (The Chinese University of Hong Kong, HKSAR), Wenlong Liao (Shanghai Jiao Tong University, China), and Junchi Yan (Shanghai Jiao Tong University, China)

NeRF-HuGS: Improved Neural Radiance Fields in Non-static Scenes Using Heuristics-Guided Segmentation 19436
Jiahao Chen (Sun Yat-sen University, China), Yipeng Qin (Cardiff University, United Kingdom), Lingjie Liu (University of Pennsylvania, USA), Jiangbo Lu (SmartMore Corporation, China), and Guanbin Li (Sun Yat-sen University, China)

Mip-Splatting: Alias-free 3D Gaussian Splatting 19447
Zehao Yu (University of Tübingen, Germany; Tübingen AI Center, Germany), Anpei Chen (University of Tübingen, Germany; Tübingen AI Center, Germany), Binbin Huang (ShanghaiTech University, China), Torsten Sattler (Czech Technical University in Prague, Czech Republic), and Andreas Geiger (University of Tübingen, Germany; Tübingen AI Center, Germany)

pixelSplat: 3D Gaussian Splats from Image Pairs for Scalable Generalizable 3D Reconstruction 19457
David Charatan (Massachusetts Institute of Technology), Sizhe Lester Li (Massachusetts Institute of Technology), Andrea Tagliasacchi (Simon Fraser University, University of Toronto), and Vincent Sitzmann (Massachusetts Institute of Technology)

Learning to Produce Semi-dense Correspondences for Visual Localization 19468
Khang Truong Giang (KAIST, Republic of Korea), Soohwan Song (Dongguk University, Republic of Korea), and Sungho Jo (KAIST, Republic of Korea)

Orals 5C Low-shot, Self-supervised, Semi-supervised Learning

CroSel: Cross Selection of Confident Pseudo Labels for Partial-Label Learning 19479
Shiyu Tian (Chongqing University, China), Hongxin Wei (Southern University of Science and Technology, China), Yiqun Wang (Chongqing University, China), and Lei Feng (Singapore University of Technology and Design, Singapore)

MLP Can Be A Good Transformer Learner 19489
Sihao Lin (RMIT University), Pumeng Lyu (Shanghai AI Laboratory), Dongrui Liu (Shanghai AI Laboratory; Shanghai Jiao Tong University), Tao Tang (Shenzhen Campus of Sun Yat-sen University), Xiaodan Liang (Shenzhen Campus of Sun Yat-sen University; DarkMatter AI Research; MBZUAI), Andy Song (RMIT University), and Xiaojun Chang (University of Technology Sydney; MBZUAI)

From SAM to CAMs: Exploring Segment Anything Model for Weakly Supervised Semantic Segmentation 19499
Hyeokjun Kweon (Korea Advanced Institute of Science and Technology, Korea) and Kuk-Jin Yoon (Korea Advanced Institute of Science and Technology, Korea)

LTGC: Long-tail Recognition via Leveraging LLMs-driven Generated Content 19510
Qihao Zhao (Beijing University of Chemical Technology), Yalun Dai (Nanyang Technological University), Hao Li (Northwestern Polytechnical University), Wei Hu (Beijing University of Chemical Technology), Fan Zhang (Beijing University of Chemical Technology), and Jun Liu (Singapore University of Technology and Design)

Improving Semantic Correspondence with Viewpoint-Guided Spherical Maps 19521
Octave Mariotti (University of Edinburgh), Oisin Mac Aodha (University of Edinburgh), and Hakan Bilen (University of Edinburgh)

Poster Session 5

TeMO: Towards Text-Driven 3D Stylization for Multi-Object Meshes 19531
Xuying Zhang (Nankai University), Bo-Wen Yin (Nankai University), Yuming Chen (Nankai University), Zheng Lin (Nankai University), Yunheng Li (Nankai University), Qibin Hou (Nankai University), and Ming-Ming Cheng (Nankai University)

Event-based Structure-from-Orbit 19541
Ethan Elms (The University of Adelaide, Australia), Yasir Latif (The University of Adelaide, Australia), Tae Ha Park (Stanford University, USA), and Tat-Jun Chin (The University of Adelaide, Australia)

Towards Large-scale 3D Representation Learning with Multi-dataset Point Prompt Training 19551
Xiaoyang Wu (The Univeristy of Hong Kong, Hong Kong), Zhuotao Tian (The Chinese University of Hong Kong, Hong Kong), Xin Wen (The Univeristy of Hong Kong, Hong Kong), Bohao Peng (The Chinese University of Hong Kong, Hong Kong), Xihui Liu (The Univeristy of Hong Kong, Hong Kong), Kaicheng Yu (Westlake University, China), and Hengshuang Zhao (The Univeristy of Hong Kong, Hong Kong)

LidaRF: Delving into Lidar for Neural Radiance Field on Street Scenes 19563
Shanlin Sun (University of California, Irvine), Bingbing Zhuang (NEC Labs America), Ziyu Jiang (NEC Labs America), Buyu Liu (NEC Labs America), Xiaohui Xie (University of California, Irvine), and Manmohan Chandraker (NEC Labs America)

Instantaneous Perception of Moving Objects in 3D 19573
Di Liu (Rutgers University), Bingbing Zhuang (NEC Labs America), Dimitris N. Metaxas (Rutgers University), and Manmohan Chandraker (NEC Labs America)

Implicit Event-RGBD Neural SLAM 19584
Delin Qu (Fudan University), Chi Yan (Shanghai AI Laboratory), Dong Wang (Shanghai AI Laboratory), Jie Yin (Shanghai Jiaotong University), Qizhi Chen (Shanghai AI Laboratory), Dan Xu (Hong Kong University of Sciences and Technology), Yiting Zhang (Shanghai AI Laboratory), Bin Zhao (Shanghai AI Laboratory), and Xuelong Li (Institute of Artificial Intelligence (TeleAI) of China Telecom)

GS-SLAM: Dense Visual SLAM with 3D Gaussian Splatting 19595
Chi Yan (Shanghai AI Laboratory), Delin Qu (Fudan University), Dan Xu (Hong Kong University of Science and Technology), Bin Zhao (Shanghai AI Laboratory), Zhigang Wang (Shanghai AI Laboratory), Dong Wang (Shanghai AI Laboratory), and Xuelong Li (Shanghai AI Laboratory)

Learning Instance-Aware Correspondences for Robust Multi-Instance Point Cloud Registration in Cluttered Scenes 19605
Zhiyuan Yu (National University of Defense Technology), Zheng Qin (National University of Defense Technology), Lintao Zheng (National University of Defense Technology), and Kai Xu (National University of Defense Technology)

MeshGPT: Generating Triangle Meshes with Decoder-Only Transformers 19615
Yawar Siddiqui (Technical University of Munich, Germany), Antonio Alliegro (Politecnico di Torino, Italy), Alexey Artemov (Technical University of Munich, Germany), Tatiana Tommasi (Politecnico di Torino, Italy), Daniele Sirigatti (AUDI AG, Germany), Vladislav Rosov (AUDI AG, Germany), Angela Dai (Technical University of Munich, Germany), and Matthias Nießner (Technical University of Munich, Germany)

Multi-Session SLAM with Differentiable Wide-Baseline Pose Optimization 19626
Lahav Lipson (Princeton University) and Jia Deng (Princeton University)

SHINOBI: Shape and Illumination using Neural Object Decomposition via BRDF Optimization In-the-wild 19636
Andreas Engelhardt (University of Tübingen, Germany), Amit Raj (Google Research, USA), Mark Boss (Stability AI, Germany), Yunzhi Zhang (Stanford University, USA), Abhishek Kar (Google Research, USA), Yuanzhen Li (Google Research, USA), Deqing Sun (Google Research, USA), Ricardo Martin Brualla (Google Research, USA), Jonathan T. Barron (Google Research, USA), Hendrik P. A. Lensch (University of Tübingen, Germany), and Varun Jampani (Stability AI, UK)

HybridNeRF: Efficient Neural Rendering via Adaptive Volumetric Surfaces 19647
Haithem Turki (Carnegie Mellon University), Vasu Agrawal (Meta Reality Labs), Samuel Rota Bulò (Meta Reality Labs), Lorenzo Porzi (Meta Reality Labs), Peter Kontschieder (Meta Reality Labs), Deva Ramanan (Carnegie Mellon University), Michael Zollhöfer (Meta Reality Labs), and Christian Richardt (Meta Reality Labs)

PLGSLAM: Progressive Neural Scene Represenation with Local to Global Bundle Adjustment ... 19657
Tianchen Deng (Shanghai Jiao Tong University), Guole Shen (Shanghai Jiao Tong University), Tong Qin (Shanghai Jiao Tong University), Jianyu Wang (Shanghai Jiao Tong University), Wentao Zhao (Shanghai Jiao Tong University), Jingchuan Wang (Shanghai Jiao Tong University), Danwei Wang (Nanyang Technological University), and Weidong Chen (Shanghai Jiao Tong University)

Gear-NeRF: Free-Viewpoint Rendering and Tracking with Motion-aware Spatio-Temporal Sampling 19667
Xinhang Liu (HKUST, Hong Kong), Yu-Wing Tai (Dartmouth College, USA), Chi-Keung Tang (HKUST, Hong Kong), Pedro Miraldo (Mitsubishi Electric Research Laboratories (MERL), USA), Suhas Lohit (Mitsubishi Electric Research Laboratories (MERL), USA), and Moitreya Chatterjee (Mitsubishi Electric Research Laboratories (MERL), USA)

GPS-Gaussian: Generalizable Pixel-wise 3D Gaussian Splatting for Real-time Human Novel View Synthesis 19680
Shunyuan Zheng (Harbin Institute of Technology, China), Boyao Zhou (Tsinghua University, China), Ruizhi Shao (Tsinghua University, China), Boning Liu (Tsinghua University, China), Shengping Zhang (Harbin Institute of Technology, China), Liqiang Nie (Harbin Institute of Technology, China), and Yebin Liu (Tsinghua University, China)

HyperSDFusion: Bridging Hierarchical Structures in Language and Geometry for Enhanced 3D Text2Shape Generation 19691
Zhiying Leng (Technical University of Munich, Germany; Beihang University, China;), Tolga Birdal (Imperial College London, U.K.), Xiaohui Liang (Beihang University, China; Zhongguancun Laboratory, China;), and Federico Tombari (Technical University of Munich, Germany)

Selective-Stereo: Adaptive Frequency Information Selection for Stereo Matching 19701
Xianqi Wang (Huazhong University of Science and Technology), Gangwei Xu (Huazhong University of Science and Technology), Hao Jia (Huazhong University of Science and Technology), and Xin Yang (Huazhong University of Science and Technology)

Animatable Gaussians: Learning Pose-dependent Gaussian Maps for High-fidelity Human Avatar Modeling 19711
Zhe Li (Tsinghua University), Zerong Zheng (NNKosmos Technology), Lizhen Wang (Tsinghua University), and Yebin Liu (Tsinghua University)

Global Latent Neural Rendering 19723
Thomas Tanay (Huawei Technologies Ltd., Noah's Ark Lab) and Matteo Maggioni (Huawei Technologies Ltd., Noah's Ark Lab)

HiFi4G: High-Fidelity Human Performance Rendering via Compact Gaussian Splatting 19734
Yuheng Jiang (ShanghaiTech University, China; NeuDim, China), Zhehao Shen (ShanghaiTech University, China), Penghao Wang (ShanghaiTech University, China), Zhuo Su (ByteDance, China), Yu Hong (ShanghaiTech University, China), Yingliang Zhang (DGene, China), Jingyi Yu (ShanghaiTech University, China), and Lan Xu (ShanghaiTech University, China)

LoS: Local Structure-Guided Stereo Matching 19746
Kunhong Li (Sun Yat-Sen University, China), Longguang Wang (Aviation University of Air Force, China), Ye Zhang (Sun Yat-Sen University, China), Kaiwen Xue (Huawei Cloud Computing Technologies Co., Ltd., China), Shunbo Zhou (Huawei Cloud Computing Technologies Co., Ltd., China), and Yulan Guo (National University of Defense Technology, China)

EmbodiedScan: A Holistic Multi-Modal 3D Perception Suite Towards Embodied AI 19757
Tai Wang (Shanghai AI Laboratory), Xiaohan Mao (Shanghai AI Laboratory, Shanghai Jiao Tong University), Chenming Zhu (Shanghai AI Laboratory, The University of Hong Kong), Runsen Xu (Shanghai AI Laboratory, The Chinese University of Hong Kong), Ruiyuan Lyu (Shanghai AI Laboratory, Tsinghua University), Peisen Li (Shanghai AI Laboratory, Tsinghua University), Xiao Chen (Shanghai AI Laboratory, The Chinese University of Hong Kong), Wenwei Zhang (Shanghai AI Laboratory), Kai Chen (Shanghai AI Laboratory), Tianfan Xue (The Chinese University of Hong Kong), Xihui Liu (The University of Hong Kong), Cewu Lu (Shanghai Jiao Tong University), Dahua Lin (Shanghai AI Laboratory, The Chinese University of Hong Kong), and Jiangmiao Pang (Shanghai AI Laboratory)

Masked Spatial Propagation Network for Sparsity-Adaptive Depth Refinement 19768
Jinyoung Jun (Korea University), Jae-Han Lee (Gauss Labs. Inc), and Chang-Su Kim (Korea University)

CausalPC: Improving the Robustness of Point Cloud Classification by Causal Effect Identification 19779
Yuanmin Huang (Fudan University), Mi Zhang (Fudan University), Daizong Ding (Fudan University), Erling Jiang (Fudan University), Zhaoxiang Wang (Fudan University), and Min Yang (Fudan University)

RoMa: Robust Dense Feature Matching 19790
Johan Edstedt (Linköping University, Sweden), Qiyu Sun (East China University of Science and Technology, China), Georg Bökman (Chalmers University of Technology, Sweden), Mårten Wadenbäck (Linköping University, Sweden), and Michael Felsberg (Linköping University, Sweden)

MVHumanNet: A Large-scale Dataset of Multi-view Daily Dressing Human Captures 19801
Zhangyang Xiong (FNii, CUHKSZ; SSE, CUHKSZ), Chenghong Li (FNii, CUHKSZ; SSE, CUHKSZ), Kenkun Liu (The Chinese University of Hong Kong, Shenzhen), Hongjie Liao (The Chinese University of Hong Kong, Shenzhen), Jianqiao Hu (The Chinese University of Hong Kong, Shenzhen), Junyi Zhu (The Chinese University of Hong Kong, Shenzhen), Shuliang Ning (The Chinese University of Hong Kong, Shenzhen), Lingteng Qiu (The Chinese University of Hong Kong, Shenzhen), Chongjie Wang (The Chinese University of Hong Kong, Shenzhen), Shijie Wang (The Chinese University of Hong Kong, Shenzhen), Shuguang Cui (SSE, CUHKSZ; FNii, CUHKSZ), and Xiaoguang Han (SSE, CUHKSZ; FNii, CUHKSZ)

GES : Generalized Exponential Splatting for Efficient Radiance Field Rendering 19812
Abdullah Hamdi (University of Oxford, United Kingdom), Luke Melas-Kyriazi (University of Oxford, United Kingdom), Jinjie Mai (KAUST, Saudi Arabia), Guocheng Qian (Snap Inc, USA), Ruoshi Liu (Columbia University, USA), Carl Vondrick (Columbia University, USA), Bernard Ghanem (KAUST, Saudi Arabia), and Andrea Vedaldi (University of Oxford, United Kingdom)

RegionPLC: Regional Point-Language Contrastive Learning for Open-World 3D Scene Understanding 19823
Jihan Yang (The University of Hong Kong), Runyu Ding (The University of Hong Kong), Weipeng Deng (The University of Hong Kong), Zhe Wang (SenseTime Research), and Xiaojuan Qi (The University of Hong Kong)

NeLF-Pro: Neural Light Field Probes for Multi-Scale Novel View Synthesis 19833
Zinuo You (ETH Zurich, Switzerland), Andreas Geiger (University of Tubingen, Germany), and Anpei Chen (University of Tubingen, Germany)

LEAP-VO: Long-term Effective Any Point Tracking for Visual Odometry 19844
Weirong Chen (TU Munich, Munich Center for Machine Learning, Germany), Le Chen (MPI for Intelligent Systems, Germany), Rui Wang (Microsoft, Switzerland), and Marc Pollefeys (Microsoft, Switzerland)

FAR: Flexible, Accurate and Robust 6DoF Relative Camera Pose Estimation 19854
Chris Rockwell (University of Michigan), Nilesh Kulkarni (University of Michigan), Linyi Jin (University of Michigan), Jeong Joon Park (University of Michigan), Justin Johnson (University of Michigan), and David F. Fouhey (New York University)

OmniGlue: Generalizable Feature Matching with Foundation Model Guidance 19865
Hanwen Jiang (UT Austin), Arjun Karpur (Google Research), Bingyi Cao (Google Research), Qixing Huang (UT Austin), and André Araujo (Google Research)

GART: Gaussian Articulated Template Models 19876
Jiahui Lei (University of Pennsylvania), Yufu Wang (University of Pennsylvania), Georgios Pavlakos (UC Berkeley), Lingjie Liu (University of Pennsylvania), and Kostas Daniilidis (University of Pennsylvania)

CG-HOI: Contact-Guided 3D Human-Object Interaction Generation 19888
Christian Diller (Technical University of Munich, Germany) and Angela Dai (Technical University of Munich, Germany)

FutureHuman3D: Forecasting Complex Long-Term 3D Human Behavior from Video Observations 19902
Christian Diller (Technical University of Munich, Germany), Thomas Funkhouser (Google, USA), and Angela Dai (Technical University of Munich, Germany)

PI3D: Efficient Text-to-3D Generation with Pseudo-Image Diffusion 19915
Ying-Tian Liu (BNRist, Tsinghua University, China), Yuan-Chen Guo (BNRist, Tsinghua University; VAST, China), Guan Luo (BNRist, Tsinghua University, China), Heyi Sun (BNRist, Tsinghua University, China), Wei Yin (The University of Adelaide, Australia), and Song-Hai Zhang (Qinghai University; BNRist, Tsinghua University, China)

Building a Strong Pre-Training Baseline for Universal 3D Large-Scale Perception 19925
Haoming Chen (East China Normal University, China), Zhizhong Zhang (East China Normal University, China), Yanyun Qu (Xiamen University, China), Ruixin Zhang (Tencent Youtu Lab, China), Xin Tan (East China Normal University, China; Chongqing Institute of East China Normal University, China), and Yuan Xie (East China Normal University, China; Chongqing Institute of East China Normal University, China)

COTR: Compact Occupancy TRansformer for Vision-based 3D Occupancy Prediction 19936
Qihang Ma (Eash China Normal University, China), Xin Tan (East China Normal University, China; Chongqing Institute of Eash China Normal University, China), Yanyun Qu (Xiamen University, China), Lizhuang Ma (East China Normal University, China), Zhizhong Zhang (East China Normal University, China), and Yuan Xie (East China Normal University, China; Chongqing Institute of East China Normal University, China)

SelfOcc: Self-Supervised Vision-Based 3D Occupancy Prediction 19946
Yuanhui Huang (Tsinghua University, China), Wenzhao Zheng (Tsinghua University, China), Borui Zhang (Tsinghua University, China), Jie Zhou (Tsinghua University, China), and Jiwen Lu (Tsinghua University, China)

UnScene3D: Unsupervised 3D Instance Segmentation for Indoor Scenes 19957
David Rozenberszki (Technical University of Munich, Germany), Or Litany (NVIDIA, USA, Technion, Israel), and Angela Dai (Technical University of Munich, Germany)

NEAT: Distilling 3D Wireframes from Neural Attraction Fields 19968
Nan Xue (Ant Group), Bin Tan (Wuhan University; Ant Group), Yuxi Xiao (Zhejiang University; Ant Group), Liang Dong (Google), Gui-Song Xia (Wuhan University), Tianfu Wu (NC State University), and Yujun Shen (Ant Group)

3DInAction: Understanding Human Actions in 3D Point Clouds 19978
Yizhak Ben-Shabat (Technionm Israel), Oren Shrout (Technion, Israel), and Stephen Gould (Australian National University, Australia)

Dynamic LiDAR Re-simulation using Compositional Neural Fields 19988
Hanfeng Wu (ETH Zurich, Switzerland; TU Munich, Germany), Xingxing Zuo (TU Munich, Germany), Stefan Leutenegger (TU Munich, Germany), Or Litany (Technion, Isarel; Nvidia, USA), Konrad Schindler (ETH Zurich, Switzerland), and Shengyu Huang (ETH Zurich, Switzerland)

Inverse Rendering of Glossy Objects via the Neural Plenoptic Function and Radiance Fields 19999
Haoyuan Wang (City University of Hong Kong, Hong Kong), Wenbo Hu (Tecent, China), Lei Zhu (City University of Hong Kong, Hong Kong), and Rynson W.H. Lau (City University of Hong Kong, Hong Kong)

PanoPose: Self-supervised Relative Pose Estimation for Panoramic Images 20009
Diantao Tu (Institute of Automation, Chinese Academy of Sciences, China), Hainan Cui (Institute of Automation, Chinese Academy of Sciences, China), Xianwei Zheng (The State Key Lab. LIESMARS, Wuhan University, China), and Shuhan Shen (Institute of Automation, Chinese Academy of Sciences, China)

GeoAuxNet: Towards Universal 3D Representation Learning for Multi-sensor Point Clouds 20019
Shengjun Zhang (Tsinghua University), Xin Fei (Tsinghua University), and Yueqi Duan (Tsinghua University)

4K4D: Real-Time 4D View Synthesis at 4K Resolution 20029
Zhen Xu (Zhejiang University, China), Sida Peng (Zhejiang University, China), Haotong Lin (Zhejiang University, China), Guangzhao He (Zhejiang University, China), Jiaming Sun (Zhejiang University, China), Yujun Shen (Ant Group, China), Hujun Bao (Zhejiang University, China), and Xiaowei Zhou (Zhejiang University, China)

MuRF: Multi-Baseline Radiance Fields 20041
Haofei Xu (ETH Zurich, University of Tübingen, Tübingen AI Center), Anpei Chen (ETH Zurich, University of Tübingen, Tübingen AI Center), Yuedong Chen (Monash University), Christos Sakaridis (ETH Zurich), Yulun Zhang (Shanghai Jiao Tong University), Marc Pollefeys (ETH Zurich, Microsoft), Andreas Geiger (University of Tübingen, Tübingen AI Center), and Fisher Yu (ETH Zurich)

LangSplat: 3D Language Gaussian Splatting 20051
Minghan Qin (Tsinghua University, China), Wanhua Li (Harvard University, USA), Jiawei Zhou (Tsinghua University, China), Haoqian Wang (Tsinghua University, China), and Hanspeter Pfister (Harvard University, USA)

Bayes' Rays: Uncertainty Quantification for Neural Radiance Fields 20061
Lily Goli (University of Toronto, Canada), Cody Reading (Simon Fraser University, Canada), Silvia Sellán (University of Toronto, Canada), Alec Jacobson (University of Toronto, Canada; Adobe Research, Canada), and Andrea Tagliasacchi (Simon Fraser University, Canada; Google DeepMind, Canada)

Accelerating Neural Field Training via Soft Mining 20071
Shakiba Kheradmand (University of British Columbia, Canada), Daniel Rebain (University of British Columbia, Canada), Gopal Sharma (University of British Columbia, Canada), Hossam Isack (Google Research, USA), Abhishek Kar (Google Research, USA), Andrea Tagliasacchi (Simon Fraser University, Canada; University of Toronto, Canada; Google DeepMind, USA), and Kwang Moo Yi (University of British Columbia, Canada)

CORE-MPI: Consistency Object Removal with Embedding MultiPlane Image 20081
Donggeun Yoon (Chungnam National University, South Korea; Korea Electronics Technology Institute (KETI), South Korea) and Donghyeon Cho (Hanyang University, South Korea)

NECA: Neural Customizable Human Avatar 20091
Junjin Xiao (Sun Yat-sen University, China), Qing Zhang (Sun Yat-sen University, China), Zhan Xu (Adobe Research, America), and Wei-Shi Zheng (Sun Yat-sen University, China)

S-DyRF: Reference-Based Stylized Radiance Fields for Dynamic Scenes 20102
Xingyi Li (Huazhong University of Science and Technology, China; Nanyang Technological University, Singapore), Zhiguo Cao (Huazhong University of Science and Technology, China), Yizheng Wu (Huazhong University of Science and Technology, China; Nanyang Technological University, Singapore), Kewei Wang (Huazhong University of Science and Technology, China; Nanyang Technological University, Singapore), Ke Xian (Huazhong University of Science and Technology, China; Nanyang Technological University, Singapore), Zhe Wang (SenseTime Research, China), and Guosheng Lin (Nanyang Technological University, Singapore)

BEVNeXt: Reviving Dense BEV Frameworks for 3D Object Detection 20113
Zhenxin Li (Fudan University, China), Shiyi Lan (NVIDIA, USA), Jose M. Alvarez (NVIDIA, USA), and Zuxuan Wu (Fudan University, China)

Bi-SSC: Geometric-Semantic Bidirectional Fusion for Camera-based 3D Semantic Scene Completion 20124
Yujie Xue (Hunan University, China), Ruihui Li (Hunan University, China), Fan Wu (Hunan University, China), Zhuo Tang (Hunan University, China), Kenli Li (Hunan University, China), and Mingxing Duan (Hunan University, China)

Learning to Select Views for Efficient Multi-View Understanding 20135
Yunzhong Hou (Australian National University, Australia), Stephen Gould (Australian National University, Australia), and Liang Zheng (Australian National University, Australia)

Outdoor Scene Extrapolation with Hierarchical Generative Cellular Automata 20145
Dongsu Zhang (Seoul National University, South Korea), Francis Williams (NVIDIA, USA), Zan Gojcic (NVIDIA, Zurich), Karsten Kreis (NVIDIA, Canada), Sanja Fidler (University of Toronto, Canada), Young Min Kim (Seoul National University, South Korea), and Amlan Kar (University of Toronto, Canada)

Spectrum AUC Difference (SAUCD): Human-aligned 3D Shape Evaluation 20155
Tianyu Luan (State University of New York at Buffalo), Zhong Li (OPPO US Research Center), Lele Chen (OPPO US Research Center), Xuan Gong (Harvard Medical School), Lichang Chen (University of Maryland, College Park), Yi Xu (OPPO US Research Center), and Junsong Yuan (State University of New York at Buffalo)

Federated Online Adaptation for Deep Stereo 20165
Matteo Poggi (University of Bologna, Italy) and Fabio Tosi (University of Bologna, Italy)

Instruct 4D-to-4D: Editing 4D Scenes as Pseudo-3D Scenes Using 2D Diffusion 20176
Linzhan Mou (Zhejiang University, China), Jun-Kun Chen (University of Illinois Urbana-Champaign, USA), and Yu-Xiong Wang (University of Illinois Urbana-Champaign, USA)

Real-time Acquisition and Reconstruction of Dynamic Volumes with Neural Structured Illumination 20186
Yixin Zeng (Zhejiang University, China), Zoubin Bi (Zhejiang University, China), Mingrui Yin (Zhejiang University, China), Xiang Feng (Zhejiang University), Kun Zhou (Zhejiang University, China), and Hongzhi Wu (Zhejiang University)

Unifying Correspondence, Pose and NeRF for Generalized Pose-Free Novel View Synthesis 20196
Sunghwan Hong (Korea University), Jaewoo Jung (Korea University), Heeseong Shin (Korea University), Jiaolong Yang (Microsoft Research Asia), Seungryong Kim (Korea University), and Chong Luo (Microsoft Research Asia)

GoMVS: Geometrically Consistent Cost Aggregation for Multi-View Stereo 20207
Jiang Wu (Northwestern Polytechnical University, China), Rui Li (Northwestern Polytechnical University, China, ETH Zürich, Switzerland), Haofei Xu (University of Tübingen, Tübingen AI Center, ETH Zürich), Wenxun Zhao (Northwestern Polytechnical University, China), Yu Zhu (Northwestern Polytechnical University, China), Jinqiu Sun (Northwestern Polytechnical University, China), and Yanning Zhang (Northwestern Polytechnical University, China)

MESA: Matching Everything by Segmenting Anything 20217
Yesheng Zhang (Shanghai Jiao Tong University, China) and Xu Zhao (Shanghai Jiao Tong University, China)

OmniSDF: Scene Reconstruction using Omnidirectional Signed Distance Functions and Adaptive Binoctrees 20227
Hakyeong Kim (KAIST), Andreas Meuleman (INRIA), Hyeonjoong Jang (KAIST), James Tompkin (Brown University), and Min H. Kim (KAIST)

MirageRoom: 3D Scene Segmentation with 2D Pre-trained Models by Mirage Projection 20237
Haowen Sun (Tsinghua University), Yueqi Duan (Tsinghua University), Juncheng Yan (Tsinghua University), Yifan Liu (Tsinghua University), and Jiwen Lu (Tsinghua University)

Robust Synthetic-to-Real Transfer for Stereo Matching 20247
Jiawei Zhang (Beihang University, China), Jiahe Li (Beihang University, China), Lei Huang (Beihang University, China), Xiaohan Yu (Macquarie University, Australia), Lin Gu (RIKEN AIP; The University of Tokyo, Japan), Jin Zheng (Beihang University, China), and Xiao Bai (Beihang University, China)

Symphonize 3D Semantic Scene Completion with Contextual Instance Queries 20258
Haoyi Jiang (Huazhong University of Science and Technology), Tianheng Cheng (Huazhong University of Science and Technology), Naiyu Gao (Horizon Robotics), Haoyang Zhang (Horizon Robotics), Tianwei Lin (Horizon Robotics), Wenyu Liu (Huazhong University of Science and Technology), and Xinggang Wang (Huazhong University of Science and Technology)

Differentiable Neural Surface Refinement for Modeling Transparent Objects 20268
Weijian Deng (The Australian National University), Dylan Campbell (The Australian National University), Chunyi Sun (The Australian National University), Shubham Kanitkar (RIOS Intelligent Machines), Matthew E. Shaffer (RIOS Intelligent Machines), and Stephen Gould (The Australian National University)

DeMatch: Deep Decomposition of Motion Field for Two-View Correspondence Learning 20278
Shihua Zhang (Wuhan University), Zizhuo Li (Wuhan University), Yuan Gao (Wuhan University), and Jiayi Ma (Wuhan University)

Is Vanilla MLP in Neural Radiance Field Enough for Few-shot View Synthesis? 20288
Hanxin Zhu (University of Science and Technology of China), Tianyu He (Microsoft Research Asia), Xin Li (University of Science and Technology of China), Bingchen Li (University of Science and Technology of China), and Zhibo Chen (University of Science and Technology of China)

GaussianAvatars: Photorealistic Head Avatars with Rigged 3D Gaussians 20299
Shenhan Qian (Technical University of Munich), Tobias Kirschstein (Technical University of Munich), Liam Schoneveld (Woven by Toyota), Davide Davoli (Technical University of Munich), Simon Giebenhain (Technical University of Munich), and Matthias Nießner (Technical University of Munich)

4D Gaussian Splatting for Real-Time Dynamic Scene Rendering 20310
Guanjun Wu (Huazhong University of Science and Technology, China), Taoran Yi (Huazhong University of Science and Technology, China), Jiemin Fang (Huawei Inc., China), Lingxi Xie (Huawei Inc., China), Xiaopeng Zhang (Huawei Inc., China), Wei Wei (Huazhong University of Science and Technology, China), Wenyu Liu (Huazhong University of Science and Technology, China), Qi Tian (Huawei Inc., China), and Xinggang Wang (Huazhong University of Science and Technology, China)

How Far Can We Compress Instant-NGP-Based NeRF? 20321
Yihang Chen (Shanghai Jiao Tong University), Qianyi Wu (Monash University), Mehrtash Harandi (Monash University), and Jianfei Cai (Monash University)

Deformable 3D Gaussians for High-Fidelity Monocular Dynamic Scene Reconstruction 20331
Ziyi Yang (Zhejiang University, China; ByteDance, China), Xinyu Gao (Zhejiang University, China), Wen Zhou (ByteDance, China), Shaohui Jiao (ByteDance, China), Yuqing Zhang (Zhejiang University, China), and Xiaogang Jin (Zhejiang University, China)

Learning with Unreliability: Fast Few-shot Voxel Radiance Fields with Relative Geometric Consistency 20342
Yingjie Xu (South China University of Technology, China; Singapore Management University, Singapore), Bangzhen Liu (South China University of Technology, China), Hao Tang (Nanjing University of Science and Technology, China; Singapore Management University, Singapore), Bailin Deng (Cardiff University, United Kingdom), and Shengfeng He (Singapore Management University, Singapore)

NTO3D: Neural Target Object 3D Reconstruction with Segment Anything 20352
Xiaobao Wei (Peking University, China), Renrui Zhang (Shanghai Artificial Intelligence Laboratory, China), Jiarui Wu (Peking University, China), Jiaming Liu (Peking University, China), Ming Lu (Intel Labs China, China), Yandong Guo (AI2 Robotics, China), and Shanghang Zhang (Peking University, China)

Loopy-SLAM: Dense Neural SLAM with Loop Closures 20363
Lorenzo Liso (ETH Zürich), Erik Sandström (ETH Zürich), Vladimir Yugay (University of Amsterdam), Luc Van Gool (ETH Zürich), and Martin R. Oswald (University of Amsterdam)

BSNet: Box-Supervised Simulation-assisted Mean Teacher for 3D Instance Segmentation 20374
Jiahao Lu (University of Science and Technology of China, China), Jiacheng Deng (University of Science and Technology of China, China), and Tianzhu Zhang (University of Science and Technology of China, China; Deep Space Exploration Lab, China)

ExtraNeRF: Visibility-Aware View Extrapolation of Neural Radiance Fields with Diffusion Models 20385
Meng-Li Shih (University of Washington), Wei-Chiu Ma (Cornell University), Lorenzo Boyice (Google), Aleksander Holynski (Google Research), Forrester Cole (Google Research), Brian Curless (University of Washington), and Janne Kontkanen (Google Research)

Alpha Invariance: On Inverse Scaling Between Distance and Volume Density in Neural Radiance Fields 20396
Joshua Ahn (University of Chicago, USA), Haochen Wang (TTI-Chicago, USA), Raymond A. Yeh (Purdue University, USA), and Greg Shakhnarovich (TTI-Chicago, USA)

SpatialTracker: Tracking Any 2D Pixels in 3D Space 20406
Yuxi Xiao (Zhejiang University), Qianqian Wang (UC Berkeley), Shangzhan Zhang (Zhejiang University), Nan Xue (Ant Group), Sida Peng (Zhejiang University), Yujun Shen (Ant Group), and Xiaowei Zhou (Zhejiang University)

GauHuman: Articulated Gaussian Splatting from Monocular Human Videos 20418
Shoukang Hu (Nanyang Technological University), Tao Hu (Nanyang Technological University), and Ziwei Liu (Nanyang Technological University)

IPoD: Implicit Field Learning with Point Diffusion for Generalizable 3D Object Reconstruction from Single RGB-D Images 20432
Yushuang Wu (The Chinese University of Hong Kong (Shenzhen), China), Luyue Shi (The Chinese University of Hong Kong (Shenzhen), China), Junhao Cai (The Hong Kong University of Science and Technology, China), Weihao Yuan (Alibaba Group, China), Lingteng Qiu (The Chinese University of Hong Kong (Shenzhen), China), Zilong Dong (Alibaba Group, China), Liefeng Bo (Alibaba Group, China), Shuguang Cui (The Chinese University of Hong Kong (Shenzhen), China), and Xiaoguang Han (The Chinese University of Hong Kong (Shenzhen), China)

GOV-NeSF: Generalizable Open-Vocabulary Neural Semantic Fields 20443
Yunsong Wang (National University of Singapore), Hanlin Chen (National University of Singapore), and Gim Hee Lee (National University of Singapore)

LASA: Instance Reconstruction from Real Scans using A Large-scale Aligned Shape Annotation Dataset 20454
Haolin Liu (The Chinese University of Hong Kong, Shenzhen), Chongjie Ye (The Chinese University of Hong Kong, Shenzhen), Yinyu Nie (Technical University of Munich), Yingfan He (The Chinese University of Hong Kong, Shenzhen), and Xiaoguang Han (The Chinese University of Hong Kong, Shenzhen)

GenZI: Zero-Shot 3D Human-Scene Interaction Generation 20465
Lei Li (Technical University of Munich, Germany) and Angela Dai (Technical University of Munich, Germany)

MVCPS-NeuS: Multi-view Constrained Photometric Stereo for Neural Surface Reconstruction ... 20475
Hiroaki Santo (Osaka University, Japan), Fumio Okura (Osaka University, Japan), and Yasuyuki Matsushita (Osaka University, Japan)

DVMNet: Computing Relative Pose for Unseen Objects Beyond Hypotheses 20485
Chen Zhao (EPFL), Tong Zhang (EPFL), Zheng Dang (EPFL), and Mathieu Salzmann (EPFL)

Motion2VecSets: 4D Latent Vector Set Diffusion for Non-rigid Shape Reconstruction and Tracking 20496
Wei Cao (Technical University of Munich), Chang Luo (Technical University of Munich), Biao Zhang (King Abdullah University of Science and Technology), Matthias Nießner (Technical University of Munich), and Jiapeng Tang (Technical University of Munich)

DiffuScene: Denoising Diffusion Models for Generative Indoor Scene Synthesis 20507
Jiapeng Tang (Technical University of Munich), Yinyu Nie (Technical University of Munich), Lev Markhasin (Sony Semiconductor Solutions Europe), Angela Dai (Technical University of Munich), Justus Thies (Technical University of Darmstadt), and Matthias Nießner (Technical University of Munich)

Test-Time Adaptation for Depth Completion 20519
Hyoungseob Park (Yale University), Anjali Gupta (Yale University), and Alex Wong (Yale University)

Global and Hierarchical Geometry Consistency Priors for Few-shot NeRFs in Indoor Scenes 20530
Xiaotian Sun (Xiamen University, China), Qingshan Xu (Nanyang Technological University, Singapore), Xinjie Yang (Xiamen University, China), Yu Zang (Xiamen University, China), and Cheng Wang (Xiamen University, China)

KP-RED: Exploiting Semantic Keypoints for Joint 3D Shape Retrieval and Deformation 20540
Ruida Zhang (Tsinghua University, China), Chenyangguang Zhang (Tsinghua University, China), Yan Di (Technical University of Munich, Germany), Fabian Manhardt (Google, Switzerland), Xingyu Liu (Tsinghua University, China), Federico Tombari (Technical University of Munich, Germany; Google, Switzerland), and Xiangyang Ji (Tsinghua University, China)

Unsigned Orthogonal Distance Fields: An Accurate Neural Implicit Representation for Diverse 3D Shapes 20551
Yujie Lu (Donghua University), Long Wan (Donghua University), Nayu Ding (Donghua University), Yulong Wang (Donghua University), Shuhan Shen (Institute of Automation, Chinese Academy of Sciences; University of Chinese Academy of Sciences), Shen Cai (Donghua University), and Lin Gao (Institute of Computing Technology, Chinese Academy of Sciences; University of Chinese Academy of Sciences)

DiSR-NeRF: Diffusion-Guided View-Consistent Super-Resolution NeRF 20561
Jie Long Lee (National University of Singapore, Singapore), Chen Li (National University of Singapore, Singapore), and Gim Hee Lee (National University of Singapore, Singapore)

BANF: Band-Limited Neural Fields for Levels of Detail Reconstruction 20571
Akhmedkhan Shabanov (Simon Fraser University), Shrisudhan Govindarajan (Simon Fraser University), Cody Reading (Simon Fraser University), Lily Goli (University of Toronto), Daniel Rebain (University of British Columbia), Kwang Moo Yi (University of British Columbia), and Andrea Tagliasacchi (Simon Fraser University,University of Toronto,Google DeepMind)

SuperNormal: Neural Surface Reconstruction via Multi-View Normal Integration 20581
Xu Cao (CyberAgent) and Takafumi Taketomi (CyberAgent)

ADFactory: An Effective Framework for Generalizing Optical Flow with NeRF 20591
Han Ling (Nanjing University of Science and Technology, China), Quansen Sun (Nanjing University of Science and Technology, China), Yinghui Sun (Southeast University, China), Xian Xu (Southeast University, China), and Xinfeng Li (Nanjing University of Science and Technology, China)

Dr.Hair: Reconstructing Scalp-Connected Hair Strands without Pre-Training via Differentiable Rendering of Line Segments 20601
Yusuke Takimoto (Huawei Technologies Japan K.K., Japan), Hikari Takehara (Huawei Technologies Japan K.K., Japan), Hiroyuki Sato (Huawei Technologies Japan K.K., Japan), Zihao Zhu (Keio University, Japan), and Bo Zheng (Huawei Technologies Japan K.K., Japan)

OmniSeg3D: Omniversal 3D Segmentation via Hierarchical Contrastive Learning 20612
Haiyang Ying (Tsinghua University), Yixuan Yin (Tsinghua University), Jinzhi Zhang (Tsinghua University), Fan Wang (Alibaba Group), Tao Yu (Tsinghua University), Ruqi Huang (Tsinghua University), and Lu Fang (Tsinghua University)

Visual Programming for Zero-shot Open-Vocabulary 3D Visual Grounding 20623
*Zhihao Yuan (CUHKSZ), Jinke Ren (CUHKSZ), Chun-Mei Feng (A*STAR, Singapore), Hengshuang Zhao (HKU), Shuguang Cui (CUHKSZ), and Zhen Li (CUHKSZ)*

GEARS: Local Geometry-aware Hand-object Interaction Synthesis 20634
Keyang Zhou (Max Planck Institute for Informatics; University of Tübingen), Bharat Lal Bhatnagar (Max Planck Institute for Informatics), Jan Eric Lenssen (Max Planck Institute for Informatics), and Gerard Pons-Moll (Max Planck Institute for Informatics; University of Tübingen)

Edge-Aware 3D Instance Segmentation Network with Intelligent Semantic Prior 20644
Wonseok Roh (Korea University, Republic of Korea), Hwanhee Jung (Korea University, Republic of Korea), Giljoo Nam (Meta Reality Labs, USA), Jinseop Yeom (Korea University, Republic of Korea), Hyunje Park (Korea University, Republic of Korea), Sang Ho Yoon (KAIST, Republic of Korea), and Sangpil Kim (Korea University, Republic of Korea)

Scaffold-GS: Structured 3D Gaussians for View-Adaptive Rendering 20654
Tao Lu (Shanghai AI Lab), Mulin Yu (Shanghai AI Lab), Linning Xu (The Chinese University of Hong Kong), Yuanbo Xiangli (Cornell University), Limin Wang (Nanjing University; Shanghai AI Lab), Dahua Lin (Shanghai AI Lab, The Chinese University of Hong Kong), and Bo Dai (Shanghai AI Lab)

Map-Relative Pose Regression for Visual Re-Localization 20665
Shuai Chen (Niantic; University of Oxford), Tommaso Cavallari (Niantic), Victor Adrian Prisacariu (Niantic; University of Oxford), and Eric Brachmann (Niantic)

3DGStream: On-the-Fly Training of 3D Gaussians for Efficient Streaming of Photo-Realistic Free-Viewpoint Videos 20675
Jiakai Sun (Zhejiang University), Han Jiao (Zhejiang University), Guangyuan Li (Zhejiang University), Zhanjie Zhang (Zhejiang University), Lei Zhao (Zhejiang University), and Wei Xing (Zhejiang University)

Revisiting Global Translation Estimation with Feature Tracks 20686
Peilin Tao (Institute of Automation, Chinese Academy of Sciences), Hainan Cui (Institute of Automation, Chinese Academy of Sciences), Mengqi Rong (Institute of Automation, Chinese Academy of Sciences), and Shuhan Shen (Institute of Automation, Chinese Academy of Sciences)

DUSt3R: Geometric 3D Vision Made Easy 20697
Shuzhe Wang (Aalto University), Vincent Leroy (Naver Labs Europe), Yohann Cabon (Naver Labs Europe), Boris Chidlovskii (Naver Labs Europe), and Jerome Revaud (Naver Labs Europe)

Robust Depth Enhancement via Polarization Prompt Fusion Tuning 20710
Kei Ikemura (KTH Royal Institute of Technology), Yiming Huang (CUHK), Felix Heide (Princeton University), Zhaoxiang Zhang (CAIR, HKISI-CAS), Qifeng Chen (HKUST), and Chenyang Lei (CAIR, HKISI-CAS)

StraightPCF: Straight Point Cloud Filtering 20721
Dasith de Silva Edirimuni (Deakin University, Australia), Xuequan Lu (La Trobe University, Australia), Gang Li (Deakin University, Australia), Lei Wei (Deakin University, Australia), Antonio Robles-Kelly (Deakin University, Australia), and Hongdong Li (Australian National University, Australia)

NeRFiller: Completing Scenes via Generative 3D Inpainting 20731
Ethan Weber (UC Berkeley), Aleksander Holynski (Google Research, UC Berkeley), Varun Jampani (Google Research), Saurabh Saxena (Google Research), Noah Snavely (Google Research), Abhishek Kar (Google Research), and Angjoo Kanazawa (UC Berkeley)

NeRF Director: Revisiting View Selection in Neural Volume Rendering 20742
Wenhui Xiao (Queensland University of Technology, Australia), Rodrigo Santa Cruz (Imaging and Computer Vision Group, CSIRO Data61, Australia), David Ahmedt-Aristizabal (Imaging and Computer Vision Group, CSIRO Data61, Australia), Olivier Salvado (Imaging and Computer Vision Group, CSIRO Data61, Australia), Clinton Fookes (Queensland University of Technology, Australia), and Leo Lebrat (Imaging and Computer Vision Group, CSIRO Data61, Australia)

Learning Intra-view and Cross-view Geometric Knowledge for Stereo Matching 20752
*Rui Gong (Nanyang Technological University), Weide Liu (Institute for Infocomm Research, A*STAR), Zaiwang Gu (Institute for Infocomm Research, A*STAR), Xulei Yang (Institute for Infocomm Research, A*STAR), and Jun Cheng (Institute for Infocomm Research, A*STAR)*

Sherpa3D: Boosting High-Fidelity Text-to-3D Generation via Coarse 3D Prior 20763
Fangfu Liu (Tsinghua University), Diankun Wu (Tsinghua University), Yi Wei (Tsinghua University), Yongming Rao (BAAI), and Yueqi Duan (Tsinghua University)

DNGaussian: Optimizing Sparse-View 3D Gaussian Radiance Fields with Global-Local Depth Normalization 20775
Jiahe Li (Beihang University, China), Jiawei Zhang (Beihang University, China), Xiao Bai (Beihang University, China), Jin Zheng (Beihang University, China), Xin Ning (Institute of Semiconductors, Chinese Academy of Sciences, China), Jun Zhou (Griffith University, Australia), and Lin Gu (RIKEN AIP / The University of Tokyo, Japan)

A Conditional Denoising Diffusion Probabilistic Model for Point Cloud Upsampling 20786
Wentao Qu (Nanjing University of Science and Technology), Yuantian Shao (Nanjing University of Science and Technology), Lingwu Meng (Nanjing University of Science and Technology), Xiaoshui Huang (Shanghai AI Laboratory), and Liang Xiao (Nanjing University of Science and Technology)

COLMAP-Free 3D Gaussian Splatting 20796
Yang Fu (University of California San Diego), Sifei Liu (Nvidia), Amey Kulkarni (Nvidia), Jan Kautz (Nvidia), Alexei A. Efros (University of California Berkeley), and Xiaolong Wang (University of California San Diego)

GSNeRF: Generalizable Semantic Neural Radiance Fields with Enhanced 3D Scene Understanding..... 20806
Zi-Ting Chou (Natioanl Taiwan University, Taiwan), Sheng-Yu Huang (Natioanl Taiwan University, Taiwan), I-Jieh Liu (Natioanl Taiwan University, Taiwan), and Yu-Chiang Frank Wang (Natioanl Taiwan University, Taiwan; NVIDIA, Taiwan)

Extend Your Own Correspondences: Unsupervised Distant Point Cloud Registration by Progressive Distance Extension 20816
Quan Liu (Shanghai Jiao Tong University), Hongzi Zhu (Shanghai Jiao Tong University), Zhenxi Wang (Shanghai Jiao Tong University), Yunsong Zhou (Shanghai Jiao Tong University), Shan Chang (Donghua University), and Minyi Guo (Shanghai Jiao Tong University)

Fully Geometric Panoramic Localization 20827
Junho Kim (Seoul National University, South Korea), Jiwon Jeong (Stanford University, USA), and Young Min Kim (Seoul National University, South Korea)

Multiway Point Cloud Mosaicking with Diffusion and Global Optimization 20838
Shengze Jin (ETH, Switzerland), Iro Armeni (Stanford, USA), Marc Pollefeys (ETH, Switzerland; Microsoft, Switzerland), and Daniel Barath (ETH, Switzerland)

Generative 3D Part Assembly via Part-Whole-Hierarchy Message Passing 20850
Bi'an Du (Peking University), Xiang Gao (Peking University), Wei Hu (Peking University), and Renjie Liao (University of British Columbia, Vector Institute for AI, Canada CIFAR AI Chair)

Total-Decom: Decomposed 3D Scene Reconstruction with Minimal Interaction 20860
Xiaoyang Lyu (The University of Hong Kong), Chirui Chang (The University of Hong Kong), Peng Dai (The University of Hong Kong), Yang-Tian Sun (The University of Hong Kong), and Xiaojuan Qi (The University of Hong Kong)

Absolute Pose from One or Two Scaled and Oriented Features 20870
Jonathan Ventura (California Polytechnic State University, USA), Zuzana Kukelova (Czech Technical University in Prague, Czech Republic), Torsten Sattler (Czech Technical University in Prague, Czech Republic), and Dániel Baráth (ETH Zürich, Switzerland)

DGC-GNN: Leveraging Geometry and Color Cues for Visual Descriptor-Free 2D-3D Matching .. 20881
Shuzhe Wang (Aalto University, Finland), Juho Kannala (Aalto University, Finland), and Daniel Barath (ETH Zurich, Switzerland)

Entity-NeRF: Detecting and Removing Moving Entities in Urban Scenes 20892
Takashi Otonari (The University of Tokyo), Satoshi Ikehata (National Institute of Informatics), and Kiyoharu Aizawa (The University of Tokyo)

GaussianEditor: Editing 3D Gaussians Delicately with Text Instructions 20902
Junjie Wang (Huawei Inc., China), Jiemin Fang (Huawei Inc., China), Xiaopeng Zhang (Huawei Inc., China), Lingxi Xie (Huawei Inc., China), and Qi Tian (Huawei Inc., China)

The More You See in 2D, the More You Perceive in 3D 20912
Xinyang Han (UC Berkeley, USA), Zelin Gao (Zhejiang University, China), Angjoo Kanazawa (UC Berkeley, USA), Shubham Goel (Avataar, USA), and Yossi Gandelsman (UC Berkeley, USA)

Multi-Scale 3D Gaussian Splatting for Anti-Aliased Rendering 20923
Zhiwen Yan (National University of Singapore), Weng Fei Low (National University of Singapore), Yu Chen (National University of Singapore), and Gim Hee Lee (National University of Singapore)

Practical Measurements of Translucent Materials with Inter-Pixel Translucency Prior 20932
Zhenyu Chen (Nanjing University, China), Jie Guo (Nanjing University, China), Shuichang Lai (Nanjing University, China), Ruoyu Fu (Nanjing University, China), Mengxun Kong (Nanjing University, China), Chen Wang (Nanjing University, China), Hongyu Sun (OPPO, USA), Zhebin Zhang (OPPO, USA), Chen Li (OPPO, USA), and Yanwen Guo (Nanjing University, China)

OneFormer3D: One Transformer for Unified Point Cloud Segmentation 20943
Maxim Kolodiazhnyi (Samsung Research, Russia), Anna Vorontsova (Samsung Research, Russia), Anton Konushin (Samsung Research, Russia), and Danila Rukhovich (Samsung Research, Russia)

General Point Model Pretraining with Autoencoding and Autoregressive 20954
Zhe Li (Huazhong University of Science and Technology), Zhangyang Gao (Westlake University), Cheng Tan (Westlake University), Bocheng Ren (Huazhong University of Science and Technology), Laurence T. Yang (Huazhong University of Science and Technology; Zhengzhou University), and Stan Z. Li (Westlake University)

MorpheuS: Neural Dynamic 360° Surface Reconstruction from Monocular RGB-D Video 20965
Hengyi Wang (University College London, United Kingdom), Jingwen Wang (University College London, United Kingdom), and Lourdes Agapito (University College London, United Kingdom)

Object Dynamics Modeling with Hierarchical Point Cloud-based Representations 20977
Chanho Kim (Oregon State University, USA) and Li Fuxin (Oregon State University, USA)

Neural Refinement for Absolute Pose Regression with Feature Synthesis 20987
Shuai Chen (University of Oxford), Yash Bhalgat (University of Oxford), Xinghui Li (University of Oxford), Jia-Wang Bian (University of Oxford), Kejie Li (University of Oxford), Zirui Wang (University of Oxford), and Victor Adrian Prisacariu (University of Oxford)

Gaussian Shadow Casting for Neural Characters 20997
Luis Bolanos (The University of British Columbia, Canada), Shih-Yang Su (The University of British Columbia, Canada), and Helge Rhodin (The University of British Columbia, Canada)

PAPR in Motion: Seamless Point-level 3D Scene Interpolation 21007
Shichong Peng (Simon Fraser University, Canada), Yanshu Zhang (Simon Fraser University), and Ke Li (Simon Fraser University)

ShapeMatcher: Self-Supervised Joint Shape Canonicalization, Segmentation, Retrieval and Deformation 21017
Yan Di (Technical University of Munich, Germany), Chenyangguang Zhang (Tsinghua University, China), Chaowei Wang (Northwestern Polytechnical University, China), Ruida Zhang (Tsinghua University, China), Guangyao Zhai (Technical University of Munich, Germany), Yanyan Li (Technical University of Munich, Germany), Bowen Fu (Tsinghua University, China), Xiangyang Ji (Tsinghua University, China), and Shan Gao (Northwestern Polytechnical University, China)

XScale-NVS: Cross-Scale Novel View Synthesis with Hash Featurized Manifold 21029
Guangyu Wang (Tsinghua University), Jinzhi Zhang (Tsinghua University), Fan Wang (Alibaba Group), Ruqi Huang (Tsinghua University), and Lu Fang (Tsinghua University)

Instance-Adaptive and Geometric-Aware Keypoint Learning for Category-Level 6D Object Pose Estimation 21040
Xiao Lin (University of Science and Technology of China), Wenfei Yang (University of Science and Technology of China), Yuan Gao (University of Science and Technology of China), and Tianzhu Zhang (University of Science and Technology of China)

RepKPU: Point Cloud Upsampling with Kernel Point Representation and Deformation 21050
Yi Rong (Nanjing University), Haoran Zhou (Nanjing University), Kang Xia (Nanjing University), Cheng Mei (Nanjing University), Jiahao Wang (Nanjing University), and Tong Lu (Nanjing University)

ColorPCR: Color Point Cloud Registration with Multi-Stage Geometric-Color Fusion 21061
Juncheng Mu (Tsinghua University, China), Lin Bie (Tsinghua University, China), Shaoyi Du (Xi'an Jiaotong University, China), and Yue Gao (Tsinghua University, China)

ConsistDreamer: 3D-Consistent 2D Diffusion for High-Fidelity Scene Editing 21071
Jun-Kun Chen (University of Illinois Urbana-Champaign, USA), Samuel Rota Bulò (Meta, Switzerland), Norman Müller (Meta, Switzerland), Lorenzo Porzi (Meta, Switzerland), Peter Kontschieder (Meta, Switzerland), and Yu-Xiong Wang (University of Illinois Urbana-Champaign, USA)

SceneTex: High-Quality Texture Synthesis for Indoor Scenes via Diffusion Priors 21081
Dave Zhenyu Chen (Technical University of Munich, Germany), Haoxuan Li (Technical University of Munich, Germany), Hsin-Ying Lee (Snap Research, USA), Sergey Tulyakov (Snap Research, USA), and Matthias Nießner (Technical University of Munich, Germany)

Aerial Lifting: Neural Urban Semantic and Building Instance Lifting from Aerial Imagery 21092
Yuqi Zhang (The Chinese University of Hongkong Shenzhen, China), Guanying Chen (The Chinese University of Hongkong Shenzhen, China; Sun Yat-sen University, China), Jiaxing Chen (The Chinese University of Hongkong Shenzhen, China; Sun Yat-sen University, China), and Shuguang Cui (The Chinese University of Hongkong Shenzhen, China)

Improving Depth Completion via Depth Feature Upsampling 21104
Yufei Wang (Northwestern Polytechnical University, China), Ge Zhang (Beijing Institute of Tracking and Telecommunication Technology, China), Shaoqian Wang (Northwestern Polytechnical University, China), Bo Li (Northwestern Polytechnical University, China), Qi Liu (Northwestern Polytechnical University, China), Le Hui (Northwestern Polytechnical University, China), and Yuchao Dai (Northwestern Polytechnical University, China)

ZeroRF: Fast Sparse View 360° Reconstruction with Zero Pretraining 21114
Ruoxi Shi (UC San Diego), Xinyue Wei (UC San Diego), Cheng Wang (UC San Diego), and Hao Su (UC San Diego)

Multi-Level Neural Scene Graphs for Dynamic Urban Environments 21125
Tobias Fischer (ETH Zurich), Lorenzo Porzi (Meta Reality Labs), Samuel Rota Bulo (Meta Reality Labs), Marc Pollefeys (ETH Zurich), and Peter Kontschieder (Meta Reality Labs)

Gaussian-Flow: 4D Reconstruction with Dynamic 3D Gaussian Particle 21136
Youtian Lin (Nanjing University, China), Zuozhuo Dai (Alibaba Group), Siyu Zhu (Fudan University), and Yao Yao (Nanjing University, China)

L4D-Track: Language-to-4D Modeling Towards 6-DoF Tracking and Shape Reconstruction in 3D Point Cloud Stream 21146
Jingtao Sun (Hunan University, China; National University of Singapore, Singapore), Yaonan Wang (Hunan University, China), Mingtao Feng (Xidian University, China; Hunan University, China), Yulan Guo (Sun Yat-Sen University), Ajmal Mian (The University of Western Australia), and Mike Zheng Shou (National University of Singapore)

Neural Directional Encoding for Efficient and Accurate View-Dependent Appearance Modeling 21157
Liwen Wu (UC San Diego), Sai Bi (Adobe Research), Zexiang Xu (Adobe Research), Fujun Luan (Adobe Research), Kai Zhang (Adobe Research), Iliyan Georgiev (Adobe Research), Kalyan Sunkavalli (Adobe Research), and Ravi Ramamoorthi (UC San Diego)

SNI-SLAM: Semantic Neural Implicit SLAM 21167
Siting Zhu (Shanghai Jiao Tong University, China), Guangming Wang (University of Cambridge, United Kingdom), Hermann Blum (ETH Zürich), Jiuming Liu (Shanghai Jiao Tong University, China), Liang Song (China University of Mining and Technology, China), Marc Pollefeys (ETH Zürich), and Hesheng Wang (Shanghai Jiao Tong University, China)

Enhancing 3D Object Detection with 2D Detection-Guided Query Anchors 21178
Haoxuanye Ji (Nullmax, China), Pengpeng Liang (Zhengzhou University, China), and Erkang Cheng (Nullmax, China)

SpecNeRF: Gaussian Directional Encoding for Specular Reflections 21188
Li Ma (The Hong Kong University of Science and Technology, Hong Kong), Vasu Agrawal (Meta Reality Labs), Haithem Turki (Carnegie Mellon University), Changil Kim (Meta Reality Labs), Chen Gao (Meta Reality Labs), Pedro Sander (The Hong Kong University of Science and Technology), Michael Zollhöfer (Meta Reality Labs), and Christian Richardt (Meta Reality Labs)

Correspondence-Free Non-Rigid Point Set Registration Using Unsupervised Clustering Analysis 21199
Mingyang Zhao (HKISI, CAS), Jingen Jiang (Shandong University), Lei Ma (Peking University), Shiqing Xin (Shandong University), Gaofeng Meng (CASIA, China), and Dong-Ming Yan (CASIA, China)

GAFusion: Adaptive Fusing LiDAR and Camera with Multiple Guidance for 3D Object Detection 21209
Xiaotian Li (Nanjing University of Posts and Telecommunications), Baojie Fan (Nanjing University of Posts and Telecommunications), Jiandong Tian (Shenyang Institute of Automation Chinese Academy of Science), and Huijie Fan (Shenyang Institute of Automation Chinese Academy of Science)

3D Neural Edge Reconstruction 21219
Lei Li (ETH Zurich, Switzerland), Songyou Peng (ETH Zurich, Switzerland), Zehao Yu (University of Tübingen, Germany), Shaohui Liu (ETH Zurich, Switzerland), Rémi Pautrat (ETH Zurich, Switzerland), Xiaochuan Yin (Utopilot, China), and Marc Pollefeys (ETH Zurich, Switzerland)

AlignMiF: Geometry-Aligned Multimodal Implicit Field for LiDAR-Camera Joint Synthesis 21230
Tang Tao (Shenzhen Campus of Sun Yat-sen University), Guangrun Wang (University of Oxford), Yixing Lao (University of Hong Kong), Peng Chen (Cainiao Group), Jie Liu (North China University of Technology), Liang Lin (Sun Yat-sen University), Kaicheng Yu (Westlake University), and Xiaodan Liang (Shenzhen Campus of Sun Yat-sen University)

Polarization Wavefront Lidar: Learning Large Scene Reconstruction from Polarized Wavefronts 21241
Dominik Scheuble (Mercedes-Benz AG, TU Darmstadt), Chenyang Lei (Princeton University), Seung-Hwan Baek (POSTECH), Mario Bijelic (Torc Robotics, Princeton University), and Felix Heide (Torc Robotics, Princeton University)

A Unified Diffusion Framework for Scene-aware Human Motion Estimation from Sparse Signals 21251
Jiangnan Tang (ShanghaiTech University), Jingya Wang (ShanghaiTech University), Kaiyang Ji (ShanghaiTech University), Lan Xu (ShanghaiTech University), Jingyi Yu (ShanghaiTech University), and Ye Shi (ShanghaiTech University)

FaceTalk: Audio-Driven Motion Diffusion for Neural Parametric Head Models 21263
Shivangi Aneja (Technical University of Munich), Justus Thies (TU Darmstadt), Angela Dai (Technical University of Munich), and Matthias Nießner (Technical University of Munich)

NeRFCodec: Neural Feature Compression Meets Neural Radiance Fields for Memory-Efficient Scene Representation 21274
Sicheng Li (Zhejiang University), Hao Li (Zhejiang University), Yiyi Liao (Zhejiang University), and Lu Yu (Zhejiang University)

Open-Vocabulary 3D Semantic Segmentation with Foundation Models 21284
Li Jiang (Max Planck Institute for Informatics, Saarland Informatics Campus), Shaoshuai Shi (Max Planck Institute for Informatics, Saarland Informatics Campus), and Bernt Schiele (Max Planck Institute for Informatics, Saarland Informatics Campus)

GraphDreamer: Compositional 3D Scene Synthesis from Scene Graphs 21295
Gege Gao (ETH Zürich, Switzerland; University of Tübingen, Germany; Tübingen AI Center, Germany; Max Planck Institute for Intelligent Systems - Tübingen, Germany), Weiyang Liu (Max Planck Institute for Intelligent Systems -- Tübingen, Germany; University of Cambridge, England), Anpei Chen (ETH Zürich, Switzerland; University of Tübingen, Germany; Tübingen AI Center, Germany), Andreas Geiger (University of Tübingen, Germany; Tübingen AI Center, Germany), and Bernhard Schölkopf (ETH Zürich, Switzerland; Tübingen AI Center, Germany; Max Planck Institute for Intelligent Systems - Tübingen, Germany)

OA-CNNs: Omni-Adaptive Sparse CNNs for 3D Semantic Segmentation 21305
Bohao Peng (The Chinese University of Hong Kong, China), Xiaoyang Wu (The University of Hong Kong, China), Li Jiang (The Chinese University of Hong Kong (Shenzhen), China), Yukang Chen (The Chinese University of Hong Kong, China), Hengshuang Zhao (The University of Hong Kong, China), Zhuotao Tian (Harbin Institute of Technology (Shenzhen), China), and Jiaya Jia (The Chinese University of Hong Kong, China)

Efficient Solution of Point-Line Absolute Pose 21316
Petr Hruby (ETH Zürich, Switzerland), Timothy Duff (University of Washington, USA), and Marc Pollefeys (ETH Zürich, Switzerland)

CN-RMA: Combined Network with Ray Marching Aggregation for 3D Indoor Object Detection from Multi-view Images 21326
Guanlin Shen (Tsinghua University, China), Jingwei Huang (Tencent, China), Zhihua Hu (Nanjing University of Information Science and Technology, China), and Bin Wang (Tsinghua University, China)

HUGS: Holistic Urban 3D Scene Understanding via Gaussian Splatting 21336
Hongyu Zhou (Zhejiang University, China), Jiahao Shao (Zhejiang University, China), Lu Xu (Zhejiang University, China), Dongfeng Bai (Huawei Noah's Ark Lab, China), Weichao Qiu (Huawei Noah's Ark Lab, China), Bingbing Liu (Huawei Noah's Ark Lab, China), Yue Wang (Zhejiang University, China), Andreas Geiger (University of Tübingen, Germany), and Yiyi Liao (Zhejiang University, China)

Benchmarking Implicit Neural Representation and Geometric Rendering in Real-Time RGB-D SLAM 21346
Tongyan Hua (Hong Kong University of Science and Technology(Guangzhou), China) and Lin Wang (Hong Kong University of Science and Technology, China)

SplaTAM: Splat, Track & Map 3D Gaussians for Dense RGB-D SLAM 21357
Nikhil Keetha (Carnegie Mellon University, USA), Jay Karhade (Carnegie Mellon University, USA), Krishna Murthy Jatavallabhula (Massachusetts Institute of Technology, USA), Gengshan Yang (Carnegie Mellon University, USA), Sebastian Scherer (Carnegie Mellon University, USA), Deva Ramanan (Carnegie Mellon University, USA), and Jonathon Luiten (Carnegie Mellon University, USA)

Lift3D: Zero-Shot Lifting of Any 2D Vision Model to 3D 21367
Mukund Varma T (University of California San Diego), Peihao Wang (University of Texas at Austin), Zhiwen Fan (University of Texas at Austin), Zhangyang Wang (University of Texas at Austin), Hao Su (University of California San Diego), and Ravi Ramamoorthi (University of California San Diego)

TutteNet: Injective 3D Deformations by Composition of 2D Mesh Deformations 21378
Bo Sun (University of Texas at Austin, USA), Thibault Groueix (Adobe Research, USA), Chen Song (University of Texas at Austin, USA), Qixing Huang (University of Texas at Austin, USA), and Noam Aigerman (University of Montreal, Canada)

L0-Sampler: An L0 Model Guided Volume Sampling for NeRF 21390
Liangchen Li (University of Science and Technology of China, China) and Juyong Zhang (University of Science and Technology of China, China)

Text-to-3D using Gaussian Splatting 21401
Zilong Chen (Tsinghua University, China), Feng Wang (Tsinghua University, China), Yikai Wang (Tsinghua University, China), and Huaping Liu (Tsinghua University, China)

TAMM: TriAdapter Multi-Modal Learning for 3D Shape Understanding 21413
Zhihao Zhang (Xi'an Jiaotong University, China), Shengcao Cao (University of Illinois at Urbana-Champaign, USA), and Yu-Xiong Wang (University of Illinois at Urbana-Champaign, USA)

FreGS: 3D Gaussian Splatting with Progressive Frequency Regularization 21424
Jiahui Zhang (Nanyang Technological University), Fangneng Zhan (Max Planck Institute for Informatics), Muyu Xu (Nanyang Technological University), Shijian Lu (Nanyang Technological University), and Eric Xing (Carnegie Mellon University)

NeISF: Neural Incident Stokes Field for Geometry and Material Estimation 21434
Chenhao Li (Osaka University), Taishi Ono (Sony Europe B.V.), Takeshi Uemori (Sony Semiconductor Solutions Corporation), Hajime Mihara (Sony Semiconductor Solutions Corporation), Alexander Gatto (Sony Europe B.V.), Hajime Nagahara (Osaka University), and Yusuke Moriuchi (Sony Semiconductor Solutions Corporation)

Non-Rigid Structure-from-Motion: Temporally-Smooth Procrustean Alignment and Spatially-Variant Deformation Modeling 21446
Jiawei Shi (Northwestern Polytechnical University, China), Hui Deng (Northwestern Polytechnical University, China), and Yuchao Dai (Northwestern Polytechnical University, China)

Small Steps and Level Sets: Fitting Neural Surface Models with Point Guidance 21456
Chamin Hewa Koneputugodage (Australian National University, Australia), Yizhak Ben-Shabat (Technion Israel Institute of Technology, Israel), Dylan Campbell (Australian National University, Australia), and Stephen Gould (Australian National University, Australia)

CVT-xRF: Contrastive In-Voxel Transformer for 3D Consistent Radiance Fields from Sparse Inputs 21466
Yingji Zhong (The Hong Kong University of Science and Technology, China), Lanqing Hong (Huawei Noah's Ark Lab, China), Zhenguo Li (Huawei Noah's Ark Lab, China), and Dan Xu (The Hong Kong University of Science and Technology, China)

GaussianEditor: Swift and Controllable 3D Editing with Gaussian Splatting 21476
Yiwen Chen (Nanyang Technological University), Zilong Chen (Tsinghua University), Chi Zhang (Nanyang Technological University), Feng Wang (Tsinghua University), Xiaofeng Yang (Nanyang Technological University), Yikai Wang (Tsinghua University), Zhongang Cai (SenseTime Research), Lei Yang (SenseTime Research), Huaping Liu (Tsinghua University), and Guosheng Lin (Nanyang Technological University)

Cam4DOcc: Benchmark for Camera-Only 4D Occupancy Forecasting in Autonomous Driving Applications 21486
Junyi Ma (Shanghai Jiao Tong University, China), Xieyuanli Chen (National University of Defense Technology, China), Jiawei Huang (HAOMO.AI, China), Jingyi Xu (Shanghai Jiao Tong University, China), Zhen Luo (Beijing Institute of Technology, China), Jintao Xu (HAOMO.AI, China), Weihao Gu (HAOMO.AI, China), Rui Ai (HAOMO.AI, China), and Hesheng Wang (Shanghai Jiao Tong University, China)

UDiFF: Generating Conditional Unsigned Distance Fields with Optimal Wavelet Diffusion 21496
Junsheng Zhou (Tsinghua University, China), Weiqi Zhang (Tsinghua University, China), Baorui Ma (BAAI, China), Kanle Shi (Kuaishou Technology, China), Yu-Shen Liu (Tsinghua University, China), and Zhizhong Han (Wayne State University, USA)

PanoRecon: Real-Time Panoptic 3D Reconstruction from Monocular Video 21507
Dong Wu (Peking University, China), Zike Yan (Tsinghua University, China), and Hongbin Zha (Peking University, China)

Three Pillars Improving Vision Foundation Model Distillation for Lidar 21519
Gilles Puy (valeo.ai, France), Spyros Gidaris (valeo.ai, France), Alexandre Boulch (valeo.ai, France), Oriane Siméoni (valeo.ai, France), Corentin Sautier (valeo.ai, France), Patrick Pérez (Kyutai, France), Andrei Bursuc (valeo.ai, France), and Renaud Marlet (valeo.ai, France)

GARField: Group Anything with Radiance Fields 21530
Chung Min Kim (UC Berkeley), Mingxuan Wu (UC Berkeley), Justin Kerr (UC Berkeley), Ken Goldberg (UC Berkeley), Matthew Tancik (Luma AI), and Angjoo Kanazawa (UC Berkeley)

Flexible Depth Completion for Sparse and Varying Point Densities 21540
Jinhyung Park (Carnegie Mellon University, USA), Yu-Jhe Li (Microsoft Research, USA), and Kris Kitani (Carnegie Mellon University, USA)

ReconFusion: 3D Reconstruction with Diffusion Priors 21551
Rundi Wu (Columbia University), Ben Mildenhall (Google Research), Philipp Henzler (Google Research), Keunhong Park (Google Research), Ruiqi Gao (Google DeepMind), Daniel Watson (Google Research), Pratul P. Srinivasan (Google Research), Dor Verbin (Google Research), Jonathan T. Barron (Google Research), Ben Poole (Google DeepMind), and Aleksander Hołyński (Google Research)

GLACE: Global Local Accelerated Coordinate Encoding 21562
Fangjinhua Wang (ETH Zurich, Switzerland), Xudong Jiang (ETH Zurich, Switzerland), Silvano Galliani (Microsoft Mixed Reality & AI Zurich Lab, Switzerland), Christoph Vogel (Microsoft Mixed Reality & AI Zurich Lab, Switzerland), and Marc Pollefeys (ETH Zurich, Switzerland)

NARUTO: Neural Active Reconstruction from Uncertain Target Observations 21572
Ziyue Feng (OPPO US Research Center, USA; Clemson University, USA), Huangying Zhan (OPPO US Research Center, USA), Zheng Chen (OPPO US Research Center, USA; Indiana University, USA), Qingan Yan (OPPO US Research Center, USA), Xiangyu Xu (OPPO US Research Center, USA), Changjiang Cai (OPPO US Research Center, USA), Bing Li (Clemson University, USA), Qilun Zhu (Clemson University, USA), and Yi Xu (OPPO US Research Center, USA)

Photo-SLAM: Real-time Simultaneous Localization and Photorealistic Mapping for Monocular, Stereo, and RGB-D Cameras 21584
Huajian Huang (The Hong Kong University of Science and Technology, Hong Kong, China), Longwei Li (Sun Yat-sen University, China), Hui Cheng (Sun Yat-sen University, China), and Sai-Kit Yeung (The Hong Kong University of Science and Technology, Hong Kong, China)

Detector-Free Structure from Motion 21594
Xingyi He (Zhejiang University), Jiaming Sun (Zhejiang University), Yifan Wang (Zhejiang University), Sida Peng (Zhejiang University), Qixing Huang (The University of Texas at Austin), Hujun Bao (Zhejiang University), and Xiaowei Zhou (Zhejiang University)

Memory-based Adapters for Online 3D Scene Perception 21604
Xiuwei Xu (Tsinghua University), Chong Xia (Tsinghua University), Ziwei Wang (CMU), Linqing Zhao (Tsinghua University), Yueqi Duan (Tsinghua University), Jie Zhou (Tsinghua University), and Jiwen Lu (Tsinghua University)

SurroundSDF: Implicit 3D Scene Understanding Based on Signed Distance Field 21614
Lizhe Liu (Xiaomi EV), Bohua Wang (Xi'an Jiaotong University), Hongwei Xie (Xiaomi EV), Daqi Liu (Xiaomi EV), Li Liu (Xiaomi EV), Zhiqiang Tian (Xi'an Jiaotong University), Kuiyuan Yang (Xiaomi EV), and Bing Wang (Xiaomi EV)

CoGS: Controllable Gaussian Splatting 21624
Heng Yu (Robotics Institute, Carnegie Mellon University), Joel Julin (Robotics Institute, Carnegie Mellon University), Zoltán A. Milacski (Robotics Institute, Carnegie Mellon University), Koichiro Niinuma (Fujitsu Research of America), and László A. Jeni (Robotics Institute, Carnegie Mellon University)

DrivingGaussian: Composite Gaussian Splatting for Surrounding Dynamic Autonomous Driving Scenes 21634
Xiaoyu Zhou (Peking University), Zhiwei Lin (Peking University), Xiaojun Shan (Peking University), Yongtao Wang (Peking University), Deqing Sun (Google Research), and Ming-Hsuan Yang (Google Research; University of California, Merced)

GS-IR: 3D Gaussian Splatting for Inverse Rendering 21644
Zhihao Liang (South China University of Technology), Qi Zhang (Tencent AI Lab), Ying Feng (Tencent AI Lab), Ying Shan (Tencent AI Lab), and Kui Jia (The Chinese University of Hong Kong, Shenzhen)

Cross-spectral Gated-RGB Stereo Depth Estimation 21654
Samuel Brucker (Torc Robotics, Germany), Stefanie Walz (Mercedes-Benz, Germany), Mario Bijelic (Princeton University, USA), and Felix Heide (Princeton University, USA)

Efficient LoFTR: Semi-Dense Local Feature Matching with Sparse-Like Speed 21666
Yifan Wang (Zhejiang University, China), Xingyi He (Zhejiang University, China), Sida Peng (Zhejiang University, China), Dongli Tan (Zhejiang University, China), and Xiaowei Zhou (Zhejiang University, China)

Feature 3DGS: Supercharging 3D Gaussian Splatting to Enable Distilled Feature Fields 21676
Shijie Zhou (University of California, Los Angeles), Haoran Chang (University of California, Los Angeles), Sicheng Jiang (University of California, Los Angeles), Zhiwen Fan (University of Texas at Austin), Zehao Zhu (University of Texas at Austin), Dejia Xu (University of Texas at Austin), Pradyumna Chari (University of California, Los Angeles), Suya You (DEVCOM Army Research Laboratory), Zhangyang Wang (University of Texas at Austin), and Achuta Kadambi (University of California, Los Angeles)

VGGSfM: Visual Geometry Grounded Deep Structure From Motion 21686
Jianyuan Wang (University of Oxford, United Kingdom), Nikita Karaev (University of Oxford, United Kingdom), Christian Rupprecht (University of Oxford, United Kingdom), and David Novotny (Meta AI, United Kingdom)

Dynamic Cues-Assisted Transformer for Robust Point Cloud Registration 21698
Hong Chen (Huazhong University of Science and Technology, China), Pei Yan (Huazhong University of Science and Technology, China), Sihe Xiang (Huazhong University of Science and Technology, China), and Yihua Tan (Huazhong University of Science and Technology, China)

GP-NeRF: Generalized Perception NeRF for Context-Aware 3D Scene Understanding 21708
Hao Li (Northwestern Polytechnology University), Dingwen Zhang (Northwestern Polytechnology University, Institute of Artificial Intelligence, Hefei Comprehensive National Science Center), Yalun Dai (Nanyang Technological University), Nian Liu (MBZUAI), Lechao Cheng (Hefei University of Technology), Jingfeng Li (Northwestern Polytechnology University), Jingdong Wang (Baidu, Inc.), and Junwei Han (Northwestern Polytechnology University)

Compact 3D Gaussian Representation for Radiance Field 21719
Joo Chan Lee (Sungkyunkwan University, South Korea), Daniel Rho (KT, South Korea), Xiangyu Sun (Sungkyunkwan University, South Korea), Jong Hwan Ko (Sungkyunkwan University, South Korea), and Eunbyung Park (Sungkyunkwan University, South Korea)

Unsupervised Occupancy Learning from Sparse Point Cloud 21729
Amine Ouasfi (Inria) and Adnane Boukhayma (Inria)

TACO: Benchmarking Generalizable Bimanual Tool-ACtion-Object Understanding 21740
Yun Liu (Tsinghua University, Shanghai Artificial Intelligence Laboratory, Shanghai Qi Zhi Institute), Haolin Yang (Beijing University of Posts and Telecommunications, China), Xu Si (Tsinghua University, China), Ling Liu (Beijing Institute of Technology, China), Zipeng Li (Tsinghua University, China), Yuxiang Zhang (Tsinghua University, China), Yebin Liu (Tsinghua University, China), and Li Yi (Tsinghua University, Shanghai Artificial Intelligence Laboratory, Shanghai Qi Zhi Institute)

ImageNet-D: Benchmarking Neural Network Robustness on Diffusion Synthetic Object 21752
Chenshuang Zhang (KAIST, South Korea), Fei Pan (University of Michigan, Ann Arbor, USA), Junmo Kim (KAIST, South Korea), In So Kweon (KAIST, South Korea), and Chengzhi Mao (McGill University, Canada and MILA, Canada)

SynFog: A Photo-realistic Synthetic Fog Dataset based on End-to-end Imaging Simulation for Advancing Real-World Defogging in Autonomous Driving 21763
Yiming Xie (Tsinghua University, China), Henglu Wei (Tsinghua University, China), Zhenyi Liu (Stanford University, USA), Xiaoyu Wang (Tsinghua University, China), and Xiangyang Ji (Tsinghua University, China)

FineSports: A Multi-person Hierarchical Sports Video Dataset for Fine-grained Action Understanding 21773
Jinglin Xu (University of Science and Technology Beijing, China), Guohao Zhao (Peking University, China), Sibo Yin (Peking University, China), Wenhao Zhou (University of Science and Technology Beijing, China), and Yuxin Peng (Peking University, China)

Infinigen Indoors: Photorealistic Indoor Scenes using Procedural Generation 21783
Alexander Raistrick (Princeton University, USA), Lingjie Mei (Princeton University, USA), Karhan Kayan (Princeton University, USA), David Yan (Princeton University, USA), Yiming Zuo (Princeton University, USA), Beining Han (Princeton University, USA), Hongyu Wen (Princeton University, USA), Meenal Parakh (Princeton University, USA), Stamatis Alexandropoulos (Princeton University, USA), Lahav Lipson (Princeton University, USA), Zeyu Ma (Princeton University, USA), and Jia Deng (Princeton University, USA)

Probing the 3D Awareness of Visual Foundation Models 21795
Mohamed El Banani (University of Michigan), Amit Raj (Google), Kevis-Kokitsi Maninis (Google), Abhishek Kar (Google), Yuanzhen Li (Google), Michael Rubinstein (Google), Deqing Sun (Google), Leonidas Guibas (Google), Justin Johnson (University of Michigan), and Varun Jampani (Google)

VBench: Comprehensive Benchmark Suite for Video Generative Models 21807
Ziqi Huang (S-Lab, Nanyang Technological University), Yinan He (Shanghai Artificial Intelligence Laboratory), Jiashuo Yu (Shanghai Artificial Intelligence Laboratory), Fan Zhang (Shanghai Artificial Intelligence Laboratory), Chenyang Si (S-Lab, Nanyang Technological University), Yuming Jiang (S-Lab, Nanyang Technological University), Yuanhan Zhang (S-Lab, Nanyang Technological University), Tianxing Wu (S-Lab, Nanyang Technological University), Qingyang Jin (S-Lab, Nanyang Technological University), Nattapol Chanpaisit (S-Lab, Nanyang Technological University), Yaohui Wang (Shanghai Artificial Intelligence Laboratory), Xinyuan Chen (Shanghai Artificial Intelligence Laboratory), Limin Wang (Nanjing University, Shanghai Artificial Intelligence Laboratory), Dahua Lin (Shanghai Artificial Intelligence Laboratory, The Chinese University of Hong Kong), Yu Qiao (Shanghai Artificial Intelligence Laboratory), and Ziwei Liu (S-Lab, Nanyang Technological University)

MAPLM: A Real-World Large-Scale Vision-Language Benchmark for Map and Traffic Scene Understanding 21819
Xu Cao (University of Illinois Urbana-Champaign, USA), Tong Zhou (Tencent T Lab, China), Yunsheng Ma (Purdue University, USA), Wenqian Ye (University of Virginia, USA), Can Cui (Purdue University, USA), Kun Tang (Tencent T Lab, China), Zhipeng Cao (Tencent T Lab, China), Kaizhao Liang (SambaNova Systems, USA), Ziran Wang (Purdue University, USA), James M. Rehg (University of Illinois Urbana-Champaign, USA), and Chao Zheng (Tencent T Lab, China)

Video Recognition in Portrait Mode 21831
Mingfei Han (University of Technology Sydney, Australia; Bytedance Inc., China; Data61 CSIRO, Australia; Mohamed bin Zayed University of Artificial Intelligence, UAE), Linjie Yang (Bytedance Inc., USA), Xiaojie Jin (Bytedance Inc., USA), Jiashi Feng (Bytedance Inc., Singapore), Xiaojun Chang (University of Technology Sydney, Australia; Mohamed bin Zayed University of Artificial Intelligence, UAE), and Heng Wang (Bytedance Inc, USA)

MMVP: A Multimodal MoCap Dataset with Vision and Pressure Sensors 21842
He Zhang (Beihang University, China), Shenghao Ren (Nanjing University, China), Haolei Yuan (Beihang University, China), Jianhui Zhao (Beihang University, China), Fan Li (Beihang University, China), Shuangpeng Sun (Tsinghua University, China), Zhenghao Liang (Beijing Weilan Technology Co., Ltd., China), Tao Yu (Tsinghua University, China), Qiu Shen (Nanjing University, China), and Xun Cao (Nanjing University, China)

What If the TV Was Off? Examining Counterfactual Reasoning Abilities of Multi-modal Language Models 21853
Letian Zhang (Tongji University), Xiaotong Zhai (University of Warwick), Zhongkai Zhao (National University of Singapore), Yongshuo Zong (University of Edinburgh), Xin Wen (The University of Hong Kong), and Bingchen Zhao (University of Edinburgh)

COCONut: Modernizing COCO Segmentation 21863
Xueqing Deng (ByteDance, USA), Qihang Yu (ByteDance, USA), Peng Wang (ByteDance, USA), Xiaohui Shen (ByteDance, USA), and Liang-Chieh Chen (ByteDance, USA)

Traffic Scene Parsing through the TSP6K Dataset 21874
Peng-Tao Jiang (Nankai University, China), Yuqi Yang (Nankai University, China), Yang Cao (Hong Kong University of Science and Technology, China), Qibin Hou (Nankai University, China), Ming-Ming Cheng (Nankai University, China), and Chunhua Shen (Zhejiang University, China)

Real Acoustic Fields: An Audio-Visual Room Acoustics Dataset and Benchmark 21886
Ziyang Chen (University of Michigan, USA), Israel D. Gebru (Codec Avatars Lab, Pittsburgh, Meta, USA), Christian Richardt (Codec Avatars Lab, Pittsburgh, Meta, USA), Anurag Kumar (Reality Labs Research, Meta, USA), William Laney (Codec Avatars Lab, Pittsburgh, Meta, USA), Andrew Owens (University of Michigan, USA), and Alexander Richard (Codec Avatars Lab, Pittsburgh, Meta, USA)

Rethinking the Evaluation Protocol of Domain Generalization 21897
Han Yu (Tsinghua University, China), Xingxuan Zhang (Tsinghua University, China), Renzhe Xu (Tsinghua University, China), Jiashuo Liu (Tsinghua University, China), Yue He (Tsinghua University, China), and Peng Cui (Tsinghua University, China)

MMSum: A Dataset for Multimodal Summarization and Thumbnail Generation of Videos 21909
Jielin Qiu (Carnegie Mellon University), Jiacheng Zhu (MIT), William Han (Carnegie Mellon University), Aditesh Kumar (Carnegie Mellon University), Karthik Mittal (Carnegie Mellon University), Claire Jin (Carnegie Mellon University), Zhengyuan Yang (Microsoft Azure AI), Linjie Li (Microsoft Azure AI), Jianfeng Wang (Microsoft Azure AI), Ding Zhao (Carnegie Mellon University), Bo Li (University of Chicago), and Lijuan Wang (Microsoft Azure AI)

Learning from Synthetic Human Group Activities 21922
Che-Jui Chang (Rutgers University), Danrui Li (Rutgers University), Deep Patel (NEC Laboratories), Parth Goel (Rutgers University), Honglu Zhou (NEC Laboratories), Seonghyeon Moon (Rutgers University), Samuel S. Sohn (Rutgers University), Sejong Yoon (The College of New Jersey), Vladimir Pavlovic (Rutgers University), and Mubbasir Kapadia (Roblox)

Instance Tracking in 3D Scenes from Egocentric Videos 21933
Yunhan Zhao (UC Irvine, USA), Haoyu Ma (UC Irvine, USA), Shu Kong (Texas A&M University, USA; Institute of Collaborative Innovation, China; University of Macau, China), and Charless Fowlkes (UC Irvine, USA)

Insect-Foundation: A Foundation Model and Large-scale 1M Dataset for Visual Insect Understanding 21945
Hoang-Quan Nguyen (University of Arkansas, USA), Thanh-Dat Truong (University of Arkansas, USA), Xuan Bac Nguyen (University of Arkansas, USA), Ashley Dowling (University of Arkansas, USA), Xin Li (The State University of New York at Albany, USA), and Khoa Luu (University of Arkansas, USA)

Low-Resource Vision Challenges for Foundation Models 21956
Yunhua Zhang (University of Amsterdam), Hazel Doughty (Leiden University), and Cees G. M. Snoek (University of Amsterdam)

OpenStreetView-5M: The Many Roads to Global Visual Geolocation 21967
Guillaume Astruc (Ecole des Ponts, France; IGN, France; CNES, France), Nicolas Dufour (Ecole des Ponts, France; Ecole Polytechnique, France), Ioannis Siglidis (Ecole des Ponts, France), Constantin Aronssohn (Ecole des Ponts, France), Nacim Bouia (Ecole des Ponts, France), Stephanie Fu (Ecole des Ponts, France; UC Berkeley, USA), Romain Loiseau (Ecole des Ponts, France; IGN, France), Van Nguyen Nguyen (Ecole des Ponts, France), Charles Raude (Ecole des Ponts, France), Elliot Vincent (Ecole des Ponts, France; Inria Paris, France), Lintao Xu (Ecole des Ponts, France), Hongyu Zhou (Ecole des Ponts, France), and Loic Landrieu (Ecole des Ponts, France)

FreeMan: Towards Benchmarking 3D Human Pose Estimation under Real-World Conditions 21978
Jiong Wang (The Chinese University of Hong Kong, Shenzhen), Fengyu Yang (The Chinese University of Hong Kong, Shenzhen), Bingliang Li (The Chinese University of Hong Kong, Shenzhen), Wenbo Gou (The Chinese University of Hong Kong, Shenzhen), Danqi Yan (The Chinese University of Hong Kong, Shenzhen), Ailing Zeng (IDEA), Yijun Gao (Tencent), Junle Wang (Tencent), Yanqing Jing (Tencent), and Ruimao Zhang (The Chinese University of Hong Kong, Shenzhen)

LiDAR-Net: A Real-scanned 3D Point Cloud Dataset for Indoor Scenes 21989
Yanwen Guo (Nanjing University), Yuanqi Li (Nanjing University), Dayong Ren (Nanjing University), Xiaohong Zhang (Nanjing University), Jiawei Li (Nanjing University), Liang Pu (Nanjing University), Changfeng Ma (Nanjing University), Xiaoyu Zhan (Nanjing University), Jie Guo (Nanjing University), Mingqiang Wei (Nanjing University of Aeronautics and Astronautics), Yan Zhang (Nanjing University), Piaopiao Yu (Nanjing University of Aeronautics and Astronautics), Shuangyu Yang (Nanjing University), Donghao Ji (Nanjing University), Huisheng Ye (Nanjing University), Hao Sun (Nanjing University), Yansong Liu (Nanjing University), Yinuo Chen (Nanjing University), Jiaqi Zhu (Nanjing University), and Hongyu Liu (Nanjing University)

View-decoupled Transformer for Person Re-identification under Aerial-ground Camera Network..... 22000
Quan Zhang (Sun Yat-Sen University, China; Johns Hopkins University, USA), Lei Wang (Sun Yat-Sen University, China), Vishal M. Patel (Johns Hopkins University, USA), Xiaohua Xie (Sun Yat-Sen University, China; Pazhou Lab (HuangPu), China; Guangdong Province Key Laboratory of Information Security Technology, China; Key Laboratory of Machine Intelligence and Advanced Computing, China), and Jianhaung Lai (Sun Yat-Sen University, China; Pazhou Lab (HuangPu), China; Guangdong Province Key Laboratory of Information Security Technology, China; Key Laboratory of Machine Intelligence and Advanced Computing, China)

UFineBench: Towards Text-based Person Retrieval with Ultra-fine Granularity 22010
Jialong Zuo (Huazhong University of Science and Technology), Hanyu Zhou (Huazhong University of Science and Technology), Ying Nie (Huawei Noah's Ark Lab), Feng Zhang (Huazhong University of Science and Technology), Tianyu Guo (Huawei Noah's Ark Lab), Nong Sang (Huazhong University of Science and Technology), Yunhe Wang (Huawei Noah's Ark Lab), and Changxin Gao (Huazhong University of Science and Technology)

Towards Automatic Power Battery Detection: New Challenge, Benchmark Dataset and Baseline 22020
Xiaoqi Zhao (Dalian University of Technology, China; X3000 Inspection Co., Ltd, China), Youwei Pang (Dalian University of Technology, China; X3000 Inspection Co., Ltd, China), Zhenyu Chen (Dalian University of Technology, China), Qian Yu (Dalian University of Technology, China), Lihe Zhang (Dalian University of Technology, China), Hanqi Liu (X3000 Inspection Co., Ltd, China), Jiaming Zuo (X3000 Inspection Co., Ltd, China), and Huchuan Lu (Dalian University of Technology, China)

Abductive Ego-View Accident Video Understanding for Safe Driving Perception 22030
Jianwu Fang (Xi'an Jiaotong University, China), Lei-lei Li (Chang'an University, China), Junfei Zhou (Chang'an University, China), Junbin Xiao (National University of Singapore, Singapore), Hongkai Yu (Cleveland State University, USA), Chen Lv (Nanyang Technological University, Singapore), Jianru Xue (Xi'an Jiaotong University, China), and Tat-Seng Chua (National University of Singapore, Singapore)

Multiagent Multitraversal Multimodal Self-Driving: Open MARS Dataset 22041
Yiming Li (New York University), Zhiheng Li (New York University), Nuo Chen (New York University), Moonjun Gong (New York University), Zonglin Lyu (New York University), Zehong Wang (New York University), Peili Jiang (New York University), and Chen Feng (New York University)

Towards Surveillance Video-and-Language Understanding: New Dataset, Baselines, and Challenges 22052
Tongtong Yuan (Beijing University of Technology, CN), Xuange Zhang (Beijing University of Technology, CN), Kun Liu (Beijing University of Posts and Telecommunications, CN), Bo Liu (Beijing University of Technology, CN), Chen Chen (University of Central Florida, USA), Jian Jin (Institute of Industrial Internet of Things, CAICT, CN), and Zhenzhen Jiao (Beijing Teleinfo Technology Co., Ltd., CAICT, CN)

Pre-training Vision Models with Mandelbulb Variations 22062
Benjamin Naoto Chiche (Rist Inc.), Yuto Horikawa (Rist Inc.), and Ryo Fujita (Rist Inc.)

EgoExoLearn: A Dataset for Bridging Asynchronous Ego- and Exo-centric View of Procedural Activities in Real World 22072
Yifei Huang (Shanghai AI Laboratory, China), Guo Chen (Shanghai AI Laboratory, China), Jilan Xu (Shanghai AI Laboratory, China), Mingfang Zhang (Shanghai AI Laboratory, China), Lijin Yang (Shanghai AI Laboratory, China), Baoqi Pei (Shanghai AI Laboratory, China), Hongjie Zhang (Shanghai AI Laboratory, China), Lu Dong (Shanghai AI Laboratory, China), Yali Wang (Shenzhen Institutes of Advanced Technology, China), Limin Wang (Nanjing University, China), and Yu Qiao (Shanghai AI Laboratory, China)

JRDB-Social: A Multifaceted Robotic Dataset for Understanding of Context and Dynamics of Human Interactions Within Social Groups 22087
Simindokht Jahangard (Monash University), Zhixi Cai (Monash University), Shiki Wen (Monash University), and Hamid Rezatofighi (Monash University)

Spectral and Polarization Vision: Spectro-polarimetric Real-world Dataset 22098
Yujin Jeon (POSTECH), Eunsue Choi (POSTECH), Youngchan Kim (POSTECH), Yunseong Moon (POSTECH), Khalid Omer (Meta), Felix Heide (Princeton University), and Seung-Hwan Baek (POSTECH)

MatSynth: A Modern PBR Materials Dataset 22109
Giuseppe Vecchio (University of Catania, Italy) and Valentin Deschaintre (Adobe Research, United Kingdom)

When Visual Grounding Meets Gigapixel-level Large-scale Scenes: Benchmark and Approach ... 22119
Tao Ma (Tsinghua Univiersity), Bing Bai (Tsinghua Univiersity), Haozhe Lin (Tsinghua Univiersity), Heyuan Wang (Peking University), Yu Wang (Tsinghua Univiersity), Lin Luo (Peking University), and Lu Fang (Tsinghua Univiersity)

HoloVIC: Large-scale Dataset and Benchmark for Multi-Sensor Holographic Intersection and Vehicle-Infrastructure Cooperative 22129
Cong Ma (SenseAuto Research), Lei Qiao (SenseAuto Research), Chengkai Zhu (SenseAuto Research), Kai Liu (SenseAuto Research), Zelong Kong (SenseAuto Research), Qing Li (SenseAuto Research), Xueqi Zhou (SenseAuto Research), Yuheng Kan (SenseAuto Research), and Wei Wu (SenseAuto Research)

EvalCrafter: Benchmarking and Evaluating Large Video Generation Models 22139
Yaofang Liu (City University of Hong Kong), Xiaodong Cun (Tencent AI Lab), Xuebo Liu (University of Macau), Xintao Wang (Tencent AI Lab), Yong Zhang (Tencent AI Lab), Haoxin Chen (Tencent AI Lab), Yang Liu (The Chinese University of Hong Kong), Tieyong Zeng (The Chinese University of Hong Kong), Raymond Chan (City University of Hong Kong), and Ying Shan (Tencent AI Lab)

Localization Is All You Evaluate: Data Leakage in Online Mapping Datasets and How to Fix It 22150
Adam Lilja (Chalmers University of Technology, Zenseact), Junsheng Fu (Zenseact), Erik Stenborg (Zenseact), and Lars Hammarstrand (Chalmers University of Technology)

DL3DV-10K: A Large-Scale Scene Dataset for Deep Learning-based 3D Vision 22160
Lu Ling (Purdue University, USA), Yichen Sheng (Purdue University, USA), Zhi Tu (Purdue University, USA), Wentian Zhao (Adobe, USA), Cheng Xin (Rutgers University, USA), Kun Wan (Adobe, USA), Lantao Yu (Adobe, USA), Qianyu Guo (Purdue University, USA), Zixun Yu (Google, USA), Yawen Lu (Purdue University, USA), Xuanmao Li (Huazhong University of Science and Technology, China), Xingpeng Sun (Purdue University, USA), Rohan Ashok (Purdue University, USA), Aniruddha Mukherjee (Purdue University, USA), Hao Kang (Wormpex AI Research, USA), Xiangrui Kong (Purdue University, USA), Gang Hua (Wormpex AI Research, USA), Tianyi Zhang (Purdue University, USA), Bedrich Benes (Purdue University, USA), and Aniket Bera (Purdue University, USA)

OmniMedVQA: A New Large-Scale Comprehensive Evaluation Benchmark for Medical LVLM . 22170
Yutao Hu (The University of Hong Kong, Hong Kong SAR; Shanghai AI Laboratory, China), Tianbin Li (Shanghai AI Laboratory, China), Quanfeng Lu (Shanghai AI Laboratory, China), Wenqi Shao (Shanghai AI Laboratory, China), Junjun He (Shanghai AI Laboratory, China), Yu Qiao (Shanghai AI Laboratory, China), and Ping Luo (The University of Hong Kong, Hong Kong SAR; Shanghai AI Laboratory, China)

Can Biases in ImageNet Models Explain Generalization? 22184
Paul Gavrikov (Offenburg University, Germany) and Janis Keuper (Offenburg University, Germany)

MVBench: A Comprehensive Multi-modal Video Understanding Benchmark 22195
Kunchang Li (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences), Yali Wang (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences), Yinan He (Shanghai AI Laboratory), Yizhuo Li (The University of Hong Kong), Yi Wang (Shanghai AI Laboratory), Yi Liu (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences), Zun Wang (Shanghai AI laboratory), Jilan Xu (Fudan University), Guo Chen (State Key Laboratory for Novel Software Technology, Nanjing University), Ping Luo (The University of Hong Kong), Limin Wang (State Key Laboratory for Novel Software Technology, Nanjing University), and Yu Qiao (Shanghai AI Laboratory)

Towards Scalable 3D Anomaly Detection and Localization: A Benchmark via 3D Anomaly Synthesis and A Self-Supervised Learning Network 22207
Wenqiao Li (ShanghaiTech University), Xiaohao Xu (University of Michigan, Ann Arbor), Yao Gu (ShanghaiTech University), Bozhong Zheng (ShanghaiTech University), Shenghua Gao (ShanghaiTech University), and Yingna Wu (ShanghaiTech University)

Point-VOS: Pointing Up Video Object Segmentation 22217
Sabarinath Mahadevan (RWTH Aachen University, Germany), Idil Esen Zulfikar (RWTH Aachen University, Germany), Paul Voigtlaender (Google Research), and Bastian Leibe (RWTH Aachen University, Germany)

GPT-4V(ision) is a Human-Aligned Evaluator for Text-to-3D Generation 22227
Tong Wu (The Chinese University of Hong Kong, Shanghai Artificial Intelligence Laboratory, Hong Kong), Guandao Yang (Stanford University, USA), Zhibing Li (The Chinese University of Hong Kong, Shanghai Artificial Intelligence Laboratory, Hong Kong), Kai Zhang (Adobe Research, USA), Ziwei Liu (S-Lab, Nanyang Technological University, Singapore), Leonidas Guibas (Stanford University, USA), Dahua Lin (The Chinese University of Hong Kong, Shanghai Artificial Intelligence Laboratory, Hong Kong), and Gordon Wetzstein (Stanford University, USA)

ConCon-Chi: Concept-Context Chimera Benchmark for Personalized Vision-Language Tasks 22239
Andrea Rosasco (University of Genoa, IT; Istituto Italiano di Tecnologia, IT), Stefano Berti (University of Genoa, IT; Istituto Italiano di Tecnologia, IT), Giulia Pasquale (Istituto Italiano di Tecnologia, IT), Damiano Malafronte (Istituto Italiano di Tecnologia, IT), Shogo Sato (Sony Interactive Entertainment Inc., JP), Hiroyuki Segawa (Sony Interactive Entertainment Inc., JP), Tetsugo Inada (Sony Interactive Entertainment Inc., JP), and Lorenzo Natale (Istituto Italiano di Tecnologia, IT)

FISBe: A Real-World Benchmark Dataset for Instance Segmentation of Long-Range Thin Filamentous Structures 22249
Lisa Mais (Max-Delbrueck-Center for Molecular Medicine in the Helmholtz Association (MDC), Germany), Peter Hirsch (Max-Delbrueck-Center for Molecular Medicine in the Helmholtz Association (MDC), Germany), Claire Managan (HHMI Janelia Research Campus, USA), Ramya Kandarpa (HHMI Janelia Research Campus, USA), Josef Lorenz Rumberger (Max-Delbrueck-Center for Molecular Medicine in the Helmholtz Association (MDC), Germany), Annika Reinke (German Cancer Research Center (DKFZ), Germany), Lena Maier-Hein (German Cancer Research Center (DKFZ), Germany), Gudrun Ihrke (HHMI Janelia Research Campus, USA), and Dagmar Kainmueller (Max-Delbrueck-Center for Molecular Medicine in the Helmholtz Association (MDC), Germany)

Inter-X: Towards Versatile Human-Human Interaction Analysis 22260
Liang Xu (Shanghai Jiao Tong University, China), Xintao Lv (Shanghai Jiao Tong University, China), Yichao Yan (Shanghai Jiao Tong University, China), Xin Jin (Eastern Institute of Technology, Ningbo, China), Shuwen Wu (Shanghai Jiao Tong University, China), Congsheng Xu (Shanghai Jiao Tong University, China), Yifan Liu (Shanghai Jiao Tong University, China), Yizhou Zhou (WeChat, Tencent Inc., China), Fengyun Rao (WeChat, Tencent Inc., China), Xingdong Sheng (Lenovo, China), Yunhui Liu (Lenovo, China), Wenjun Zeng (Eastern Institute of Technology, Ningbo, China), and Xiaokang Yang (Shanghai Jiao Tong University, China)

TextNeRF: A Novel Scene-Text Image Synthesis Method based on Neural Radiance Fields 22272
Jialei Cui (Peking University, China), Jianwei Du (Southeast University, China), Wenzhuo Liu (Peking University, China), and Zhouhui Lian (Peking University, China)

Systematic Comparison of Semi-supervised and Self-supervised Learning for Medical Image Classification 22282
Zhe Huang (Tufts University, USA), Ruijie Jiang (Tufts University), Shuchin Aeron (Tufts University, USA), and Michael C. Hughes (Tufts University)

Unexplored Faces of Robustness and Out-of-Distribution: Covariate Shifts in Environment and Sensor Domains 22294
Eunsu Baek (Seoul National University, Republic of Korea), Keondo Park (Seoul National University, Republic of Korea), Jiyoon Kim (Seoul National University, Republic of Korea), and Hyung-Sin Kim (Seoul National University, Republic of Korea)

MCD: Diverse Large-Scale Multi-Campus Dataset for Robot Perception 22304
Thien-Minh Nguyen (Nanyang Technological University), Shenghai Yuan (Nanyang Technological University), Thien Hoang Nguyen (Nanyang Technological University), Pengyu Yin (Nanyang Technological University), Haozhi Cao (Nanyang Technological University), Lihua Xie (Nanyang Technological University), Maciej Wozniak (KTH Royal Institute of Technology), Patric Jensfelt (KTH Royal Institute of Technology), Marko Thiel (The Hamburg University of Technology), Justin Ziegenbein (The Hamburg University of Technology), and Noel Blunder (The Hamburg University of Technology)

360Loc: A Dataset and Benchmark for Omnidirectional Visual Localization with Cross-device Queries 22314
Huajian Huang (The Hong Kong University of Science and Technology, Hong Kong, China), Changkun Liu (The Hong Kong University of Science and Technology, Hong Kong, China), Yipeng Zhu (The Hong Kong University of Science and Technology, Hong Kong, China), Hui Cheng (Sun Yat-sen University, China), Tristan Braud (The Hong Kong University of Science and Technology, Hong Kong, China), and Sai-Kit Yeung (The Hong Kong University of Science and Technology, Hong Kong, China)

JRDB-PanoTrack: An Open-world Panoptic Segmentation and Tracking Robotic Dataset in Crowded Human Environments 22325
Duy Tho Le (Monash University, Australia), Chenhui Gou (Monash University, Australia), Stavya Datta (Monash University, Australia), Hengcan Shi (Monash University, Australia), Ian Reid (Mohamed bin Zayed University of Artificial Intelligence, United Arab Emirates), Jianfei Cai (Monash University, Australia), and Hamid Rezatofighi (Monash University, Australia)

MTMMC: A Large-Scale Real-World Multi-Modal Camera Tracking Benchmark 22335
Sanghyun Woo (New York University), Kwanyong Park (ETRI), Inkyu Shin (KAIST), Myungchul Kim (KAIST), and In So Kweon (KAIST)

RCooper: A Real-world Large-scale Dataset for Roadside Cooperative Perception 22347
Ruiyang Hao (Tsinghua University, China), Siqi Fan (Tsinghua University, China), Yingru Dai (Tsinghua University, China), Zhenlin Zhang (China Automotive Innovation Corporation, China), Chenxi Li (China Automotive Innovation Corporation, China), Yuntian Wang (China Automotive Innovation Corporation, China), Haibao Yu (Tsinghua University, China), Wenxian Yang (Tsinghua University, China), Jirui Yuan (Tsinghua University, China), and Zaiqing Nie (Tsinghua University, China)

UVEB: A Large-scale Benchmark and Baseline Towards Real-World Underwater Video Enhancement. 22358
Yaofeng Xie (Ocean University of China), Lingwei Kong (Ocean University of China), Kai Chen (Ocean University of China), Ziqiang Zheng (The Hong Kong University of Science and Technology), Xiao Yu (Ocean University of China), Zhibin Yu (Ocean University of China), and Bing Zheng (Ocean University of China)

Real-World Mobile Image Denoising Dataset with Efficient Baselines 22368
Roman Flepp (ETH Zürich, Switzerland), Andrey Ignatov (ETH Zürich, AI Witchlabs Ltd, Switzerland), Radu Timofte (ETH Zürich, AI Witchlabs Ltd, Switzerland, University of Würzburg, Germany), and Luc Van Gool (ETH Zürich, AI Witchlabs Ltd, Switzerland)

RGBD Objects in the Wild: Scaling Real-World 3D Object Learning from RGB-D Videos 22378
Hongchi Xia (Shanghai Jiao Tong University, China), Yang Fu (University of California San Diego, USA), Sifei Liu (NVIDIA, USA), and Xiaolong Wang (University of California San Diego, USA)

Evaluating Transferability in Retrieval Tasks: An Approach Using MMD and Kernel Methods ... 22390
Mengyu Dai (Microsoft), Amir Hossein Raffiee (Salesforce), Aashish Jain (Salesforce), and Joshua Correa (Salesforce)

BEHAVIOR Vision Suite: Customizable Dataset Generation via Simulation 22401
Yunhao Ge (Stanford University, USA; University of Southern California, USA), Yihe Tang (Stanford University, USA), Jiashu Xu (Harvard University), Cem Gokmen (Stanford University, USA), Chengshu Li (Stanford University, USA), Wensi Ai (Stanford University, USA), Benjamin Jose Martinez (Stanford University, USA), Arman Aydin (Stanford University, USA), Mona Anvari (Stanford University, USA), Ayush K Chakravarthy (Stanford University, USA), Hong-Xing Yu (Stanford University, USA), Josiah Wong (Stanford University, USA), Sanjana Srivastava (Stanford University, USA), Sharon Lee (Stanford University, USA), Shengxin Zha (Meta, USA), Laurent Itti (University of Southern California, USA), Yunzhu Li (Stanford University, USA; The University of Illinois Urbana-Champaign, USA), Roberto Martín-Martín (The University of Texas at Austin, USA), Miao Liu (Meta, USA), Pengchuan Zhang (Meta, USA), Ruohan Zhang (Stanford University, USA), Li Fei-Fei (Stanford University, USA), and Jiajun Wu (Stanford University, USA)

MULAN: A Multi Layer Annotated Dataset for Controllable Text-to-Image Generation 22413
Petru-Daniel Tudosiu (Huawei Noah's Ark Lab, United Kingdom), Yongxin Yang (Huawei Noah's Ark Lab, United Kingdom), Shifeng Zhang (Huawei Noah's Ark Lab, China), Fei Chen (Huawei Noah's Ark Lab, China), Steven McDonagh (University of Edinburgh, United Kingdom), Gerasimos Lampouras (Huawei Noah's Ark Lab, United Kingdom), Ignacio Iacobacci (Huawei Noah's Ark Lab, United Kingdom), and Sarah Parisot (Huawei Noah's Ark Lab, United Kingdom)

Sieve: Multimodal Dataset Pruning using Image Captioning Models 22423
Anas Mahmoud (University of Toronto, Canada), Mostafa Elhoushi (FAIR at Meta, USA), Amro Abbas (FAIR at Meta, USA), Yu Yang (UC Los Angeles, USA), Newsha Ardalani (FAIR at Meta, USA), Hugh Leather (FAIR at Meta, USA), and Ari S. Morcos (DatologyAI, USA)

Perceptual Assessment and Optimization of HDR Image Rendering 22433
Peibei Cao (City University of Hong Kong), Rafal K. Mantiuk (University of Cambridge), and Kede Ma (City University of Hong Kong)

GlitchBench: Can Large Multimodal Models Detect Video Game Glitches? 22444
Mohammad Reza Taesiri (University of Alberta), Tianjun Feng (University of Alberta), Cor-Paul Bezemer (University of Alberta), and Anh Nguyen (Auburn University)

WinSyn: : A High Resolution Testbed for Synthetic Data 22456
Tom Kelly (King Abdullah University of Science and Technology: Thuwal, SA), John Femiani (Miami University, OH), and Peter Wonka (King Abdullah University of Science and Technology: Thuwal, SA)

DiVa-360: The Dynamic Visual Dataset for Immersive Neural Fields 22466
Cheng-You Lu (Brown University, USA), Peisen Zhou (Brown University, USA), Angela Xing (Brown University, USA), Chandradeep Pokhariya (IIIT Hyderabad, India), Arnab Dey (I3S-CNRS/Université Côte d'Azur, France), Ishaan Nikhil Shah (IIIT Hyderabad, India), Rugved Mavidipalli (Brown University, USA), Dylan Hu (Brown University, USA), Andrew I. Comport (I3S-CNRS/Université Côte d'Azur, France), Kefan Chen (Brown University, USA), and Srinath Sridhar (Brown University, USA)

Learning Discriminative Dynamics with Label Corruption for Noisy Label Detection 22477
Suyeon Kim (POSTECH), Dongha Lee (Yonsei University), SeongKu Kang (University of Illinois at Urbana Champaign), Sukang Chae (POSTECH), Sanghwan Jang (POSTECH), and Hwanjo Yu (POSTECH)

DriveTrack: A Benchmark for Long-Range Point Tracking in Real-World Videos 22488
Arjun Balasingam (MIT CSAIL), Joseph Chandler (MIT CSAIL), Chenning Li (MIT CSAIL), Zhoutong Zhang (Adobe Systems), and Hari Balakrishnan (MIT CSAIL)

HouseCat6D - A Large-Scale Multi-Modal Category Level 6D Object Perception Dataset with Household Objects in Realistic Scenarios 22498
HyunJun Jung (Technical University of Munich), Shun-Cheng Wu (Technical University of Munich), Patrick Ruhkamp (Technical University of Munich), Guangyao Zhai (Technical University of Munich, Munich Center for Machine Learning), Hannah Schieber (Technical University of Munich, FAU Erlangen-Nuernberg), Giulia Rizzoli (University of Padova), Pengyuan Wang (Technical University of Munich), Hongcheng Zhao (Technical University of Munich), Lorenzo Garattoni (Toyota Motor Europe), Sven Meier (Toyota Motor Europe), Daniel Roth (Technical University of Munich), Nassir Navab (Technical University of Munich), and Benjamin Busam (Technical University of Munich, Munich Center for Machine Learning, 3dwe.ai)

Benchmarking Segmentation Models with Mask-Preserved Attribute Editing 22509
Zijin Yin (Beijing University of Posts and Telecommunications), Kongming Liang (Beijing University of Posts and Telecommunications), Bing Li (King Abdullah University of Science and Technology), Zhanyu Ma (Beijing University of Posts and Telecommunications), and Jun Guo (Beijing University of Posts and Telecommunications)

The Devil is in the Fine-Grained Details: Evaluating Open-Vocabulary Object Detectors for Fine-Grained Understanding 22520
Lorenzo Bianchi (CNR-ISTI), Fabio Carrara (CNR-ISTI), Nicola Messina (CNR-ISTI), Claudio Gennaro (CNR-ISTI), and Fabrizio Falchi (CNR-ISTI)

PKU-DyMVHumans: A Multi-View Video Benchmark for High-Fidelity Dynamic Human Modeling .. 22530
Xiaoyun Zheng (Peking University Shenzhen Graduate School, China), Liwei Liao (Peking University Shenzhen Graduate School, China), Xufeng Li (City University of Hong Kong, China), Jianbo Jiao (University of Birmingham, United Kingdom), Rongjie Wang (Peng Cheng Laboratory, China), Feng Gao (Peking University, China), Shiqi Wang (City University of Hong Kong, China), and Ronggang Wang (Peking University Shenzhen Graduate School, China)

Insights from the Use of Previously Unseen Neural Architecture Search Datasets 22541
Rob Geada (Red Hat), David Towers (Newcastle University), Matthew Forshaw (Newcastle University), Amir Atapour-Abarghouei (Durham University), and A. Stephen McGough (Newcastle University)

TULIP: Multi-camera 3D Precision Assessment of Parkinson's Disease 22551
Kyungdo Kim (Duke University, USA), Sihan Lyu (Duke University, USA), Sneha Mantri (Duke University, USA), and Timothy W. Dunn (Duke University, USA)

LUWA Dataset: Learning Lithic Use-Wear Analysis on Microscopic Images 22563
Jing Zhang (New York University), Irving Fang (New York University), Hao Wu (New York University), Akshat Kaushik (New York University), Alice Rodriguez (New York University), Hanwen Zhao (New York University), Juexiao Zhang (New York University), Zhuo Zheng (Stanford University), Radu Iovita (New York University), and Chen Feng (New York University)

ShapeWalk: Compositional Shape Editing Through Language-Guided Chains 22574
Habib Slim (KAUST) and Mohamed Elhoseiny (KAUST)

TRINS: Towards Multimodal Language Models that Can Read 22584
Ruiyi Zhang (Adobe Research), Yanzhe Zhang (Georgia Institute of Technology), Jian Chen (State University of New York at Buffalo), Yufan Zhou (Adobe Research), Jiuxiang Gu (Adobe Research), Changyou Chen (State University of New York at Buffalo), and Tong Sun (Adobe Research)

MAGICK: A Large-scale Captioned Dataset from Matting Generated Images using Chroma Keying 22595
Ryan D. Burgert (Stony Brook University), Brian L. Price (Adobe Research), Jason Kuen (Adobe Research), Yijun Li (Adobe Research), and Michael S. Ryoo (Stony Brook University)

EFHQ: Multi-purpose ExtremePose-Face-HQ dataset 22605
Trung Tuan Dao (VinAI Research, Vietnam), Duc Hong Vu (VinAI Research, Vietnam), Cuong Pham (VinAI Research, Vietnam), and Anh Tran (VinAI Research, Vietnam)

How to Train Neural Field Representations: A Comprehensive Study and Benchmark 22616
Samuele Papa (University of Amsterdam, Netherlands), Riccardo Valperga (University of Amsterdam, Netherlands), David Knigge (University of Amsterdam, Netherlands), Miltiadis Kofinas (University of Amsterdam, Netherlands), Phillip Lippe (University of Amsterdam, Netherlands), Jan-Jakob Sonke (The Netherlands Cancer Institute, Netherlands), and Efstratios Gavves (University of Amsterdam, Netherlands)

A Noisy Elephant in the Room: Is Your Out-of-Distribution Detector Robust to Label Noise? 22626
Galadrielle Humblot-Renaux (Aalborg University, Denmark), Sergio Escalera (Aalborg University, Denmark; University of Barcelona, Spain; Computer Vision Center, Spain), and Thomas B. Moeslund (Aalborg University, Denmark)

eTraM: Event-based Traffic Monitoring Dataset 22637
Aayush Atul Verma (Arizona State University), Bharatesh Chakravarthi (Arizona State University), Arpitsinh Vaghela (Arizona State University), Hua Wei (Arizona State University), and Yezhou Yang (Arizona State University)

SubT-MRS Dataset: Pushing SLAM Towards All-weather Environments 22647
Shibo Zhao (Carnegie Mellon University, USA), Yuanjun Gao (Carnegie Mellon University, USA), Tianhao Wu (Carnegie Mellon University, USA), Damanpreet Singh (Carnegie Mellon University, USA), Rushan Jiang (Carnegie Mellon University, USA), Haoxiang Sun (Carnegie Mellon University, USA), Mansi Sarawata (Carnegie Mellon University, USA), Yuheng Qiu (Carnegie Mellon University, USA), Warren Whittaker (Carnegie Mellon University, USA), Ian Higgins (Carnegie Mellon University, USA), Yi Du (University at Buffalo, USA), Shaoshu Su (University at Buffalo, USA), Can Xu (Carnegie Mellon University, USA), John Keller (Carnegie Mellon University, USA), Jay Karhade (Carnegie Mellon University, USA), Lucas Nogueira (Carnegie Mellon University, USA), Sourojit Saha (Carnegie Mellon University, USA), Ji Zhang (Carnegie Mellon University, USA), Wenshan Wang (Carnegie Mellon University, USA), Chen Wang (University at Buffalo, USA), and Sebastian Scherer (Carnegie Mellon University, USA)

MSU-4S - The Michigan State University Four Seasons Dataset 22658
Daniel Kent (Michigan State University), Mohammed Alyaqoub (Michigan State University), Xiaohu Lu (Michigan State University), Hamed Khatounabadi (Michigan State University), Kookjin Sung (Michigan State University), Cole Scheller (Michigan State University), Alexander Dalat (Michigan State University), Asma bin Thabit (Michigan State University), Roberto Whitley (Michigan State University), and Hayder Radha (Michigan State University)

TUMTraf V2X Cooperative Perception Dataset 22668
Walter Zimmer (Technical University of Munich, Germany), Gerhard Arya Wardana (Technical University of Munich, Germany), Suren Sritharan (Technical University of Munich, Germany), Xingcheng Zhou (Technical University of Munich, Germany), Rui Song (Fraunhofer Institute for Transportation and Infrastructure Systems, Germany), and Alois C. Knoll (Technical University of Munich, Germany)

Multiview Aerial Visual RECognition (MAVREC): Can Multi-view Improve Aerial Visual Perception? 22678
Aritra Dutta (University of Central Florida), Srijan Das (University of North Carolina at Charlotte), Jacob Nielsen (University of Southern Denmark), Rajatsubhra Chakraborty (University of North Carolina at Charlotte), and Mubarak Shah (University of Central Florida)

Towards Co-Evaluation of Cameras, HDR, and Algorithms for Industrial-Grade 6DoF Pose Estimation 22691
Agastya Kalra (Intrinsic Innovation LLC, USA), Guy Stoppi (Intrinsic Innovation LLC, USA), Dmitrii Marin (Intrinsic Innovation LLC, USA), Vage Taamazyan (Intrinsic Innovation LLC, USA), Aarrushi Shandilya (Intrinsic Innovation LLC, USA), Rishav Agarwal (Intrinsic Innovation LLC, USA), Anton Boykov (Intrinsic Innovation LLC, USA), Tze Hao Chong (Intrinsic Innovation LLC, USA), and Michael Stark (Intrinsic Innovation LLC, USA)

Scaling Laws for Data Filtering— Data Curation cannot be Compute Agnostic 22702
Sachin Goyal (Carnegie Mellon University, USA), Pratyush Maini (Carnegie Mellon University, USA), Zachary C. Lipton (Carnegie Mellon University, USA), Aditi Raghunathan (Carnegie Mellon University, USA), and J. Zico Kolter (Carnegie Mellon University, USA)

Benchmarking Audio Visual Segmentation for Long-Untrimmed Videos 22712
Chen Liu (The University of Queensland), Peike Patrick Li (Matrix Verse), Qingtao Yu (Australian National University), Hongwei Sheng (The University of Queensland), Dadong Wang (CSIRO DATA61), Lincheng Li (NetEase Fuxi AI Lab), and Xin Yu (The University of Queensland)

Domain-Specific Block Selection and Paired-View Pseudo-Labeling for Online Test-Time Adaptation 22723
Yeonguk Yu (Gwangju Institute of Science and Technology, South Korea), Sungho Shin (Gwangju Institute of Science and Technology, South Korea), Seunghyeok Back (Gwangju Institute of Science and Technology, South Korea), Mihwan Ko (Gwangju Institute of Science and Technology, South Korea), Sangjun Noh (Gwangju Institute of Science and Technology, South Korea), and Kyoobin Lee (Gwangju Institute of Science and Technology, South Korea)

VideoMAC: Video Masked Autoencoders Meet ConvNets 22733
Gensheng Pei (Nanjing University of Science and Technology, China), Tao Chen (Nanjing University of Science and Technology, China), Xiruo Jiang (Nanjing University of Science and Technology, China), Huafeng Liu (Nanjing University of Science and Technology, China), Zeren Sun (Nanjing University of Science and Technology, China), and Yazhou Yao (Nanjing University of Science and Technology, China)

Unsupervised Universal Image Segmentation 22744
Dantong Niu (UC Berkeley), Xudong Wang (UC Berkeley), Xinyang Han (UC Berkeley), Long Lian (UC Berkeley), Roei Herzig (UC Berkeley), and Trevor Darrell (UC Berkeley)

VideoCutLER: Surprisingly Simple Unsupervised Video Instance Segmentation 22755
Xudong Wang (UC Berkeley, USA), Ishan Misra (Meta AI, USA), Ziyun Zeng (UC Berkeley, USA), Rohit Girdhar (Meta AI, USA), and Trevor Darrell (UC Berkeley, USA)

What You See is What You GAN: Rendering Every Pixel for High-Fidelity Geometry in 3D GANs 22765
Alex Trevithick (University of California, San Diego, USA), Matthew Chan (NVIDIA, USA), Towaki Takikawa (NVIDIA, USA), Umar Iqbal (NVIDIA, USA), Shalini De Mello (NVIDIA, USA), Manmohan Chandraker (University of California, San Diego, USA), Ravi Ramamoorthi (University of California, San Diego, USA), and Koki Nagano (NVIDIA, USA)

SPOT: Self-Training with Patch-Order Permutation for Object-Centric Learning with Autoregressive Transformers 22776
Ioannis Kakogeorgiou (National Technical University of Athens), Spyros Gidaris (valeo.ai), Konstantinos Karantzalos (National Technical University of Athens), and Nikos Komodakis (University of Crete)

Unsupervised Learning of Category-Level 3D Pose from Object-Centric Videos 22787
Leonhard Sommer (University of Freiburg), Artur Jesslen (University of Freiburg), Eddy Ilg (Saarland University), and Adam Kortylewski (University of Freiburg)

Distributionally Generative Augmentation for Fair Facial Attribute Classification 22797
Fengda Zhang (Zhejiang University, China), Qianpei He (Zhejiang University, China), Kun Kuang (Zhejiang University, China), Jiashuo Liu (Tsinghua University, China), Long Chen (Hong Kong University of Science and Technology, China), Chao Wu (Zhejiang University, China), Jun Xiao (Zhejiang University, China), and Hanwang Zhang (Nanyang Technological University, Singapore)

Estimating Noisy Class Posterior with Part-level Labels for Noisy Label Learning 22809
Rui Zhao (Xi'an Jiaotong University, China), Bin Shi (Xi'an Jiaotong University, China), Jianfei Ruan (Xi'an Jiaotong University, China), Tianze Pan (Xi'an Jiaotong University, China), and Bo Dong (Xi'an Jiaotong University, China)

Unsupervised Keypoints from Pretrained Diffusion Models 22820
Eric Hedlin (University of British Columbia, Canada), Gopal Sharma (University of British Columbia, Canada), Shweta Mahajan (University of British Columbia, Canada; Vector Institute for AI, Canada), Xingzhe He (University of British Columbia, Canada), Hossam Isack (Google Research, USA), Abhishek Kar (Google Research, USA), Helge Rhodin (University of British Columbia, Canada), Andrea Tagliasacchi (Google DeepMind, USA; Simon Fraser University, Canada; University of Toronto, Canada), and Kwang Moo Yi (University of British Columbia, Canada)

Learning to Rank Patches for Unbiased Image Redundancy Reduction 22831
Yang Luo (Fudan University), Zhineng Chen (Fudan University), Peng Zhou (University of Maryland, College Park), Zuxuan Wu (Fudan University), Xieping Gao (Hunan Normal University), and Yu-Gang Jiang (Fudan University)

Rethinking the Representation in Federated Unsupervised Learning with Non-IID Data 22841
Xinting Liao (Zhejiang University), Weiming Liu (Zhejiang University), Chaochao Chen (Zhejiang University), Pengyang Zhou (Zhejiang University), Fengyuan Yu (Zhejiang University), Huabin Zhu (Zhejiang University), Binhui Yao (Zhejiang University, Midea Group), Tao Wang (Midea Group), Xiaolin Zheng (Zhejiang University), and Yanchao Tan (Fuzhou University)

GLID: Pre-training a Generalist Encoder-Decoder Vision Model 22851
Jihao Liu (Chinese University of Hong Kong), Jinliang Zheng (Institute for AI Industry Research (AIR), Tsinghua University), Yu Liu (SenseTime Research), and Hongsheng Li (Chinese University of Hong Kong)

Sequential Modeling Enables Scalable Learning for Large Vision Models 22861
Yutong Bai (Johns Hopkins University), Xinyang Geng (UC Berkeley), Karttikeya Mangalam (UC Berkeley), Amir Bar (UC Berkeley), Alan L. Yuille (Johns Hopkins University), Trevor Darrell (UC Berkeley), Jitendra Malik (UC Berkeley), and Alexei A. Efros (UC Berkeley)

VoCo: A Simple-yet-Effective Volume Contrastive Learning Framework for 3D Medical Image Analysis 22873
Linshan Wu (Hong Kong University of Science and Technology), Jiaxin Zhuang (Hong Kong University of Science and Technology), and Hao Chen (Hong Kong University of Science and Technology)

Real-IAD: A Real-World Multi-View Dataset for Benchmarking Versatile Industrial Anomaly Detection 22883
Chengjie Wang (Shanghai Jiao Tong University; Tencent Youtu Lab), Wenbing Zhu (Fudan University; Rongcheer Co. Ltd), Bin-Bin Gao (Tencent Youtu Lab), Zhenye Gan (Tencent Youtu Lab), Jiangning Zhang (Tencent Youtu Lab), Zhihao Gu (Shanghai Jiao Tong University), Shuguang Qian (Roogcheer Co., Ltd), Mingang Chen (Shanghai Development Center of Computer Software Technology), and Lizhuang Ma (Shanghai Jiao Tong University)

BEM: Balanced and Entropy-based Mix for Long-Tailed Semi-Supervised Learning 22893
Hongwei Zheng (Meituan, China; Shanghai Jiao Tong University, China), Linyuan Zhou (Meituan, China), Han Li (Shanghai Jiao Tong University, China), Jinming Su (Meituan, China), Xiaoming Wei (Meituan, China), and Xiaoming Xu (Meituan, China)

ReCoRe: Regularized Contrastive Representation Learning of World Model 22904
Rudra P.K. Poudel (Toshiba Europe Ltd, UK), Harit Pandya (Toshiba Europe Ltd, UK), Stephan Liwicki (Toshiba Europe Ltd, UK), and Roberto Cipolla (University of Cambridge, UK)

Universal Novelty Detection Through Adaptive Contrastive Learning 22914
Hossein Mirzaei (Sharif University of Technology, Iran), Mojtaba Nafez (Sharif University of Technology, Iran), Mohammad Jafari (Sharif University of Technology, Iran), Mohammad Bagher Soltani (Sharif University of Technology, Iran), Mohammad Azizmalayeri (Sharif University of Technology, Iran), Jafar Habibi (Sharif University of Technology, Iran), Mohammad Sabokrou (Okinawa Institute of Science and Technology, Japan), and Mohammad Hossein Rohban (Sharif University of Technology, Iran)

Learning to Count without Annotations 22924
Lukas Knobel (University of Amsterdam), Tengda Han (University of Oxford), and Yuki M. Asano (University of Amsterdam)

Point Cloud Pre-training with Diffusion Models 22935
Xiao Zheng (Shandong University, China), Xiaoshui Huang (Shanghai AI Laboratory, China), Guofeng Mei (Fondazione Bruno Kessler, Italy), Yuenan Hou (Shanghai AI Laboratory, China), Zhaoyang Lyu (Shanghai AI Laboratory, China), Bo Dai (Shanghai AI Laboratory, China), Wanli Ouyang (Shanghai AI Laboratory, China), and Yongshun Gong (Shandong University, China)

Improving Unsupervised Hierarchical Representation with Reinforcement Learning 22946
Ruyi An (Nanyang Technological University), Yewen Li (Nanyang Technological University), Xu He (Huawei Noah's Ark Lab), Pengjie Gu (Nanyang Technological University), Mengchen Zhao (South China University of Technology), Dong Li (Huawei Noah's Ark Lab), Jianye Hao (Tianjin University), Chaojie Wang (Nanyang Technological University), Bo An (Nanyang Technological University), and Mingyuan Zhou (The University of Texas at Austin)

Investigating and Mitigating the Side Effects of Noisy Views for Self-Supervised Clustering Algorithms in Practical Multi-View Scenarios 22957
Jie Xu (University of Electronic Science and Technology of China, China), Yazhou Ren (University of Electronic Science and Technology of China, China; University of Electronic Science and Technology of China, China), Xiaolong Wang (University of Electronic Science and Technology of China, China), Lei Feng (Singapore University of Technology and Design, Singapore), Zheng Zhang (Harbin Institute of Technology, China), Gang Niu (RIKEN Center for Advanced Intelligence Project, Japan), and Xiaofeng Zhu (University of Electronic Science and Technology of China, China; University of Electronic Science and Technology of China, China)

Self-Supervised Representation Learning from Arbitrary Scenarios 22967
Zhaowen Li (Foundation Model Research Center, Institute of Automation, Chinese Academy of Science), Yousong Zhu (Foundation Model Research Center, Institute of Automation, Chinese Academy of Science), Zhiyang Chen (Foundation Model Research Center, Institute of Automation, Chinese Academy of Science), Zongxin Gao (Independent Researcher), Rui Zhao (Qing Yuan Research Institute, Shanghai Jiao Tong University), Chaoyang Zhao (Foundation Model Research Center, Institute of Automation, Chinese Academy of Science), Ming Tang (Foundation Model Research Center, Institute of Automation, Chinese Academy of Science), and Jinqiao Wang (Foundation Model Research Center, Institute of Automation, Chinese Academy of Science)

Learning SO(3)-Invariant Semantic Correspondence via Local Shape Transform 22978
Chunghyun Park (POSTECH, Republic of Korea), Seungwook Kim (POSTECH, Republic of Korea), Jaesik Park (Seoul National University, Republic of Korea), and Minsu Cho (POSTECH, Republic of Korea)

A Bayesian Approach to OOD Robustness in Image Classification 22988
Prakhar Kaushik (Johns Hopkins University), Adam Kortylewski (University of Freiburg), and Alan Yuille (Johns Hopkins University)

Sculpting Holistic 3D Representation in Contrastive Language-Image-3D Pre-training 22998
Yipeng Gao (Sun Yat-sen University, China), Zeyu Wang (University of California, Santa Cruz, USA), Wei-Shi Zheng (Sun Yat-sen University, China), Cihang Xie (University of California, Santa Cruz, USA), and Yuyin Zhou (University of California, Santa Cruz, USA)

Solving Masked Jigsaw Puzzles with Diffusion Vision Transformers 23009
Jinyang Liu (Northeastern University, USA), Wondmgezahu Teshome (Northeastern University, USA), Sandesh Ghimire (Qualcomm, USA), Mario Sznaier (Northeastern University, USA), and Octavia Camps (Northeastern University, USA)

DS-NeRV: Implicit Neural Video Representation with Decomposed Static and Dynamic Codes .. 23019
Hao Yan (Tianjin University, China), Zhihui Ke (Tianjin University, China), Xiaobo Zhou (Tianjin University, China), Tie Qiu (Tianjin University, China), Xidong Shi (Tianjin University, China), and Dadong Jiang (Tianjin University, China)

Brain Decodes Deep Nets 23030
Huzheng Yang (University of Pennsylvania), James Gee (University of Pennsylvania), and Jianbo Shi (University of Pennsylvania)

Pose-Guided Self-Training with Two-Stage Clustering for Unsupervised Landmark Discovery .. 23041
Siddharth Tourani (University of Heidelberg, Germany; Mohammad bin Zayed University of Artificial Intelligence, UAE), Ahmed Alwheibi (Mohamed bin Zayed University of Artificial Intelligence), Arif Mahmood (Information Technology University of Punjab), and Muhammad Haris Khan (Mohamed bin Zayed University of Artificial Intelligence)

Mitigating Object Dependencies: Improving Point Cloud Self-Supervised Learning through Object Exchange 23052
Yanhao Wu (Xi'an Jiaotong University, China), Tong Zhang (EPFL Switzerland), Wei Ke (Xi'an Jiaotong University, China), Congpei Qiu (Xi'an Jiaotong University, China), Sabine Süsstrunk (EPFL Switzerland), and Mathieu Salzmann (EPFL Switzerland)

Adaptive Slot Attention: Object Discovery with Dynamic Slot Number 23062
Ke Fan (Fudan University), Zechen Bai (National University of Singapore), Tianjun Xiao (Amazon Web Services), Tong He (Amazon Web Services), Max Horn (GSK.ai), Yanwei Fu (Fudan University), Francesco Locatello (Institute of Science and Technology Austria), and Zheng Zhang (Amazon Web Services)

Targeted Representation Alignment for Open-World Semi-Supervised Learning 23072
Ruixuan Xiao (Zhejiang University, China), Lei Feng (Singapore University of Technology and Design, Singapore), Kai Tang (Zhejiang University, China), Junbo Zhao (Zhejiang University, China), Yixuan Li (University of Wisconsin-Madison, USA), Gang Chen (Zhejiang University, China), and Haobo Wang (Zhejiang University, China)

Hierarchical Correlation Clustering and Tree Preserving Embedding 23083
Morteza Haghir Chehreghani (Chalmers University of Technology) and Mostafa Haghir Chehreghani (Amirkabir University of Technology (Tehran Polytechnic))

Contrastive Mean-Shift Learning for Generalized Category Discovery 23094
Sua Choi (Pohang University of Science and Technology (POSTECH), South Korea), Dahyun Kang (Pohang University of Science and Technology (POSTECH), South Korea), and Minsu Cho (Pohang University of Science and Technology (POSTECH), South Korea)

CuVLER: Enhanced Unsupervised Object Discoveries through Exhaustive Self-Supervised Transformers 23105
Shahaf Arica (Technion - Israel Institute of Technology), Or Rubin (Technion - Israel Institute of Technology), Sapir Gershov (Technion - Israel Institute of Technology), and Shlomi Laufer (Technion - Israel Institute of Technology)

SODA: Bottleneck Diffusion Models for Representation Learning 23115
Drew A. Hudson (Google DeepMind, UK), Daniel Zoran (Google DeepMind, UK), Mateusz Malinowski (Google DeepMind, UK), Andrew K. Lampinen (Google DeepMind, UK), Andrew Jaegle (Google DeepMind, UK), James L. McClelland (Google DeepMind, UK), Loic Matthey (Google DeepMind, UK), Felix Hill (Google DeepMind, UK), and Alexander Lerchner (Google DeepMind, UK)

HPL-ESS: Hybrid Pseudo-Labeling for Unsupervised Event-based Semantic Segmentation 23128
Linglin Jing (Shanghai AI Laboratory; Loughborough University), Yiming Ding (Shanghai AI Laboratory), Yunpeng Gao (Shanghai AI Laboratory; Northwestern Polytechnical University), Zhigang Wang (Shanghai AI Laboratory), Xu Yan (CUHK-Shenzhen), Dong Wang (Shanghai AI Laboratory), Gerald Schaefer (Loughborough University), Hui Fang (Loughborough University), Bin Zhao (Shanghai AI Laboratory; Northwestern Polytechnical University), and Xuelong Li (Shanghai AI Laboratory; Institute of Artificial Intelligence (TeleAI))

Positive-Unlabeled Learning by Latent Group-Aware Meta Disambiguation 23138
Lin Long (Zhejiang University), Haobo Wang (Zhejiang University), Zhijie Jiang (Zhejiang University), Lei Feng (Singapore University of Technology and Design), Chang Yao (Zhejiang University), Gang Chen (Zhejiang University), and Junbo Zhao (Zhejiang University)

Aligning Logits Generatively for Principled Black-Box Knowledge Distillation 23148
Jing Ma (Huazhong University of Science and Technology, China), Xiang Xiang (Huazhong University of Science and Technology, China), Ke Wang (Alibaba Damo Academy Hangzhou, China), Yuchuan Wu (Alibaba Damo Academy Hangzhou, China), and Yongbin Li (Alibaba Damo Academy Hangzhou, China)

Neural Modes: Self-supervised Learning of Nonlinear Modal Subspaces 23158
Jiahong Wang (ETH Zürich), Yinwei Du (ETH Zürich), Stelian Coros (ETH Zürich), and Bernhard Thomaszewski (ETH Zürich)

Decentralized Directed Collaboration for Personalized Federated Learning 23168
Yingqi Liu (Nanjing University of Science and Technology, China), Yifan Shi (Tsinghua University, China), Qinglun Li (National University of Defense Technology, China), Baoyuan Wu (The Chinese University of Hong Kong, China), Xueqian Wang (Tsinghua University, China), and Li Shen (JD Explore Academy, China)

Improving Graph Contrastive Learning via Adaptive Positive Sampling 23179
Jiaming Zhuo (Hebei University of Technology), Feiyang Qin (Hebei University of Technology), Can Cui (Hebei University of Technology), Kun Fu (Hebei University of Technology), Bingxin Niu (Hebei University of Technology), Mengzhu Wang (Hebei University of Technology), Yuanfang Guo (Beihang University), Chuan Wang (Institute of Information Engineering Chinese Academy of Sciences), Zhen Wang (Northwestern Polytechnical University), Xiaochun Cao (Sun Yat-sen University), and Liang Yang (Hebei University of Technology)

Integrating Efficient Optimal Transport and Functional Maps For Unsupervised Shape Correspondence Learning 23188
Tung Le (University of California, Irvine, USA), Khai Nguyen (The University of Texas at Austin, USA), Shanlin Sun (University of California, Irvine, USA), Nhat Ho (The University of Texas at Austin), and Xiaohui Xie (University of California, Irvine, USA)

Unsupervised Feature Learning with Emergent Data-Driven Prototypicality 23199
Yunhui Guo (University of Texas at Dallas), Youren Zhang (University of Michigan), Yubei Chen (University of California, Davis), and Stella X. Yu (University of Michigan)

Label Propagation for Zero-shot Classification with Vision-Language Models 23209
Vladan Stojnić (Czech Technical University in Prague), Yannis Kalantidis (NAVER LABS Europe), and Giorgos Tolias (Czech Technical University in Prague)

Boosting Continual Learning of Vision-Language Models via Mixture-of-Experts Adapters 23219
Jiazuo Yu (Dalian University of Technology, China), Yunzhi Zhuge (Dalian University of Technology, China), Lu Zhang (Dalian University of Technology, China), Ping Hu (University of Electronic Science and Technology of China), Dong Wang (Dalian University of Technology, China), Huchuan Lu (Dalian University of Technology, China), and You He (Tsinghua University, China)

Backpropagation-free Network for 3D Test-time Adaptation 23231
Yanshuo Wang (Australian National University, Australia), Ali Cheraghian (Data61-CSIRO, Australia), Zeeshan Hayder (Data61-CSIRO, Australia), Jie Hong (Australian National University, Australia), Sameera Ramasinghe (Amazon, Australia), Shafin Rahman (North South University, Bangladesh), David Ahmedt-Aristizabal (Data61-CSIRO, Australia), Xuesong Li (Australian National University, Australia), Lars Petersson (Data61-CSIRO, Australia), and Mehrtash Harandi (Monash University, Australia)

GDA: Generalized Diffusion for Robust Test-time Adaptation 23242
Yun-Yun Tsai (Columbia University, USA), Fu-Chen Chen (Amazon, USA), Albert Y. C. Chen (Amazon, USA), Junfeng Yang (Columbia University, USA), Che-Chun Su (Amazon, USA), Min Sun (Amazon, USA), and Cheng-Hao Kuo (Amazon, USA)

Semantically-Shifted Incremental Adapter-Tuning is A Continual ViTransformer 23252
Yuwen Tan (Huazhong University of Science and Technology, China), Qinhao Zhou (Huazhong University of Science and Technology, China), Xiang Xiang (Huazhong University of Science and Technology, China), Ke Wang (Alibaba Group, China), Yuchuan Wu (Alibaba Group, China), and Yongbin Li (Alibaba Group, China)

Few-shot Learner Parameterization by Diffusion Time-steps 23263
Zhongqi Yue (Nanyang Technological University), Pan Zhou (Singapore Management University), Richang Hong (Hefei University of Technology), Hanwang Zhang (Skywork AI), and Qianru Sun (Singapore Management University)

FREE: Faster and Better Data-Free Meta-Learning 23273
Yongxian Wei (Tsinghua Shenzhen International Graduate School, China), Zixuan Hu (Tsinghua Shenzhen International Graduate School, China), Zhenyi Wang (University of Maryland, College Park, USA), Li Shen (JD Explore Academy, China), Chun Yuan (Tsinghua Shenzhen International Graduate School, China), and Dacheng Tao (Nanyang Technological University, Singapore)

Classes Are Not Equal: An Empirical Study on Image Recognition Fairness 23283
Jiequan Cui (Nanyang Technological University, Singapore), Beier Zhu (Nanyang Technological University, Singapore), Xin Wen (The University of Hong Kong, Hong Kong), Xiaojuan Qi (The University of Hong Kong, Hong Kong), Bei Yu (The Chinese University of Hong Kong, Hong Kong), and Hanwang Zhang (Nanyang Technological University, Singapore)

DAVE - A Detect-and-Verify Paradigm for Low-Shot Counting 23293
Jer Pelhan (University of Ljubljana, Slovenia), Alan Lukežič (University of Ljubljana, Slovenia), Vitjan Zavrtanik (University of Ljubljana, Slovenia), and Matej Kristan (University of Ljubljana, Slovenia)

Density-guided Translator Boosts Synthetic-to-Real Unsupervised Domain Adaptive Segmentation of 3D Point Clouds 23303
Zhimin Yuan (Xiamen University, China), Wankang Zeng (Xiamen University, China), Yanfei Su (Xiamen University, China), Weiquan Liu (Xiamen University, China), Ming Cheng (Xiamen University, China), Yulan Guo (National University of Defense Technology, China), and Cheng Wang (Xiamen University, China)

D3T: Distinctive Dual-Domain Teacher Zigzagging Across RGB-Thermal Gap for Domain-Adaptive Object Detection 23313
Dinh Phat Do (Ajou University), Taehoon Kim (Ajou University), Jaemin Na (Tech. Innovation Group, KT), Jiwon Kim (Hyundai Motor Company), Keonho Lee (Hyundai Motor Company), Kyunghwan Cho (Hyundai Motor Company), and Wonjun Hwang (Ajou University)

AMU-Tuning: Effective Logit Bias for CLIP-based Few-shot Learning 23323
Yuwei Tang (Tianjin University, China), Zhenyi Lin (Tianjin University, China), Qilong Wang (Tianjin University, China), Pengfei Zhu (Tianjin University, China), and Qinghua Hu (Tianjin University, China)

LEAD: Learning Decomposition for Source-free Universal Domain Adaptation 23334
Sanqing Qu (Tongji University, China), Tianpei Zou (Tongji University, China), Lianghua He (Tongji University, China), Florian Röhrbein (Chemnitz University of Technology, Germany), Alois Knoll (Technical University of Munich, Germany), Guang Chen (Tongji University, China), and Changjun Jiang (Tongji University, China)

Improving Generalized Zero-Shot Learning by Exploring the Diverse Semantics from External Class Names 23344
Yapeng Li (Wuhan University, China), Yong Luo (Wuhan University, China), Zengmao Wang (Wuhan University, China), and Bo Du (Wuhan University, China)

What, How, and When Should Object Detectors Update in Continually Changing Test Domains? 23354
Jayeon Yoo (Seoul National University), Dongkwan Lee (Seoul National University), Inseop Chung (Seoul National University), Donghyun Kim (Korea University), and Nojun Kwak (Seoul National University)

Split to Merge: Unifying Separated Modalities for Unsupervised Domain Adaptation 23364
Xinyao Li (University of Electronic Science and Technology of China), Yuke Li (Boston College), Zhekai Du (University of Electronic Science and Technology of China), Fengling Li (University of Technology Sydney), Ke Lu (University of Electronic Science and Technology of China), and Jingjing Li (University of Electronic Science and Technology of China)

Domain-Agnostic Mutual Prompting for Unsupervised Domain Adaptation 23375
Zhekai Du (University of Electronic Science and Technology of China), Xinyao Li (University of Electronic Science and Technology of China), Fengling Li (University of Technology Sydney), Ke Lu (University of Electronic Science and Technology of China), Lei Zhu (Tongji University), and Jingjing Li (University of Electronic Science and Technology of China)

Improving the Generalization of Segmentation Foundation Model under Distribution Shift via Weakly Supervised Adaptation 23385
Haojie Zhang (South China University of Technology, China), Yongyi Su (South China University of Technology, China), Xun Xu (Institute for Infocomm Research, Singapore), and Kui Jia (The Chinese University of Hong Kong, China)

DeiT-LT: Distillation Strikes Back for Vision Transformer Training on Long-Tailed Datasets..... 23396
Harsh Rangwani (Indian Institute of Science, Bangalore), Pradipto Mondal (Indian Institute of Technology, Kharagpur), Mayank Mishra (Indian Institute of Science, Bangalore), Ashish Ramayee Asokan (Indian Institute of Science, Bangalore), and R. Venkatesh Babu (Indian Institute of Science, Bangalore)

Unified Language-driven Zero-shot Domain Adaptation 23407
Senqiao Yang (The Chinese University of Hong Kong), Zhuotao Tian (Harbin Institute of Technology, Shenzhen), Li Jiang (The Chinese University of Hong Kong, Shenzhen), and Jiaya Jia (The Chinese University of Hong Kong)

Stable Neighbor Denoising for Source-free Domain Adaptive Segmentation 23416
Dong Zhao (Xidian University), Shuang Wang (Xidian University), Qi Zang (Xidian University), Licheng Jiao (Xidian University, Shaanxi, China), Nicu Sebe (University of Trento), and Zhun Zhong (University of Nottingham)

A Simple Recipe for Language-guided Domain Generalized Segmentation 23428
Mohammad Fahes (Inria, France), Tuan-Hung Vu (Valeo.ai, France), Andrei Bursuc (Valeo.ai, France), Patrick Pérez (Kyutai, France), and Raoul de Charette (Inria, France)

TCP:Textual-based Class-aware Prompt tuning for Visual-Language Model 23438
Hantao Yao (Institute of Automation, Chinese Academy of Sciences, China), Rui Zhang (Institute of Computing Technology, Chinese Academy of Sciences, China), and Changsheng Xu (Institute of Automation, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China)

Adapters Strike Back 23449
Jan-Martin O. Steitz (TU Darmstadt) and Stefan Roth (TU Darmstadt, hessian.AI)

Improving Plasticity in Online Continual Learning via Collaborative Learning 23460
Maorong Wang (The University of Tokyo, Japan), Nicolas Michel (Univ Gustave Eiffel, France), Ling Xiao (The University of Tokyo, Japan), and Toshihiko Yamasaki (The University of Tokyo, Japan)

Visual Prompting for Generalized Few-shot Segmentation: A Multi-scale Approach 23470
Mir Rayat Imtiaz Hossain (University of British Columbia, Canada), Mennatullah Siam (Ontario Tech University, Canada), Leonid Sigal (University of British Columbia, Canada), and James J. Little (University of British Columbia, Canada)

Adaptive Random Feature Regularization on Fine-tuning Deep Neural Networks 23481
Shin'ya Yamaguchi (NTT / Kyoto University), Sekitoshi Kanai (NTT), Kazuki Adachi (NTT), and Daiki Chijiwa (NTT)

ESCAPE: Encoding Super-keypoints for Category-Agnostic Pose Estimation 23491
Khoi Duc Nguyen (University of Wisconsin-Madison), Chen Li (National University of Singapore), and Gim Hee Lee (National University of Singapore)

PracticalDG: Perturbation Distillation on Vision-Language Models for Hybrid Domain Generalization 23501
Zining Chen (Beijing University of Posts and Telecommunications), Weiqiu Wang (Beijing University of Posts and Telecommunications), Zhicheng Zhao (Beijing University of Posts and Telecommunications), Fei Su (Beijing University of Posts and Telecommunications), Aidong Men (Beijing University of Posts and Telecommunications), and Hongying Meng (Brunel University Uxbridge)

Rethinking Multi-domain Generalization with A General Learning Objective 23512
Zhaorui Tan (Xi'an Jiaotong-Liverpool University), Xi Yang (Xi'an Jiaotong-Liverpool University), and Kaizhu Huang (Duke Kunshan University)

L2B: Learning to Bootstrap Robust Models for Combating Label Noise 23523
Yuyin Zhou (University of California, Santa Cruz), Xianhang Li (University of California, Santa Cruz), Fengze Liu (Johns Hopkins University), Qingyue Wei (Stanford University), Xuxi Chen (UT Austin), Lequan Yu (The University of Hong Kong), Cihang Xie (University of California, Santa Cruz), Matthew P. Lungren (Stanford University), and Lei Xing (Stanford University)

Meta-Point Learning and Refining for Category-Agnostic Pose Estimation 23534
Junjie Chen (Jiangxi University of Finance and Economics), Jiebin Yan (Jiangxi University of Finance and Economics), Yuming Fang (Jiangxi University of Finance and Economics), and Li Niu (Shanghai Jiao Tong University)

A2XP: Towards Private Domain Generalization 23544
Geunhyeok Yu (Kyung Hee University, Republic of Korea) and Hyoseok Hwang (Kyung Hee University, Republic of Korea)

Expandable Subspace Ensemble for Pre-Trained Model-Based Class-Incremental Learning 23554
Da-Wei Zhou (Nanjing University), Hai-Long Sun (Nanjing University), Han-Jia Ye (Nanjing University), and De-Chuan Zhan (Nanjing University)

VRP-SAM: SAM with Visual Reference Prompt 23565
Yanpeng Sun (Nanjing University of Science and Technology), Jiahui Chen (Beihang University), Shan Zhang (Australian National University), Xinyu Zhang (Baidu VIS), Qiang Chen (Baidu VIS), Gang Zhang (Baidu VIS), Errui Ding (Baidu VIS), Jingdong Wang (Baidu VIS), and Zechao Li (Nanjing University of Science and Technology)

Flatten Long-Range Loss Landscapes for Cross-Domain Few-Shot Learning 23575
Yixiong Zou (Huazhong University of Science and Technology), Yicong Liu (Huazhong University of Science and Technology), Yiman Hu (Huazhong University of Science and Technology), Yuhua Li (Huazhong University of Science and Technology), and Ruixuan Li (Huazhong University of Science and Technology)

MAP: MAsk-Pruning for Source-Free Model Intellectual Property Protection 23585
Boyang Peng (Tongji University, China), Sanqing Qu (Tongji University, China), Yong Wu (Tongji University), Tianpei Zou (Tongji University, China), Lianghua He (Tongji University, China), Alois Knoll (Technical University of Munich, Germany), Guang Chen (Tongji University, China), and Changjun Jiang (Tongji University, China)

Disentangled Prompt Representation for Domain Generalization 23595
De Cheng (Xidian University, China), Zhipeng Xu (Xidian University, China), Xinyang Jiang (Microsoft Research Asia, China), Nannan Wang (Xidian University), Dongsheng Li (Microsoft Research Asia, China), and Xinbo Gao (Chongqing University of Posts and Telecommunications, China)

Adapt Before Comparison: A New Perspective on Cross-Domain Few-Shot Segmentation 23605
Jonas Herzog (Zhejiang University)

Convolutional Prompting meets Language Models for Continual Learning 23616
Anurag Roy (Indian Institute of Technology, Kharagpur), Riddhiman Moulick (Indian Institute of Technology, Kharagpur), Vinay K. Verma (Amazon India), Saptarshi Ghosh (Indian Institute of Technology, Kharagpur), and Abir Das (Indian Institute of Technology, Kharagpur)

Visual-Augmented Dynamic Semantic Prototype for Generative Zero-Shot Learning 23627
Wenjin Hou (Huazhong University of Science and Technology), Shiming Chen (Mohamed bin Zayed University of Artificial Intelligence), Shuhuang Chen (Huazhong University of Science and Technology), Ziming Hong (University of Sydney), Yan Wang (Alibaba Group), Xuetao Feng (Alibaba Group), Salman Khan (Mohamed bin Zayed University of Artificial Intelligence), Fahad Shahbaz Khan (Linköping University), and Xinge You (Huazhong University of Science and Technology)

InfLoRA: Interference-Free Low-Rank Adaptation for Continual Learning 23638
Yan-Shuo Liang (Nanjing University) and Wu-Jun Li (Nanjing University)

Discriminative Pattern Calibration Mechanism for Source-Free Domain Adaptation 23648
Haifeng Xia (Southeast University, China), Siyu Xia (Southeast University, China), and Zhengming Ding (Tulane University, USA)

NICE: Neurogenesis Inspired Contextual Encoding for Replay-free Class Incremental Learning.. 23659
Mustafa Burak Gurbuz (Georgia Institute of Technology, USA), Jean Michael Moorman (Georgia Institute of Technology, USA), and Constantine Dovrolis (The Cyprus Institute, Cyprus; Georgia Institute of Technology, USA)

Orchestrate Latent Expertise: Advancing Online Continual Learning with Multi-Level Supervision and Reverse Self-Distillation 23670
Hongwei Yan (Tsinghua University), Liyuan Wang (Tsinghua University), Kaisheng Ma (Tsinghua University), and Yi Zhong (Tsinghua University)

A Closer Look at the Few-Shot Adaptation of Large Vision-Language Models 23681
Julio Silva-Rodríguez (ÉTS Montreal, Canada), Sina Hajimiri (ÉTS Montreal, Canada), Ismail Ben Ayed (ÉTS Montreal, Canada), and Jose Dolz (ÉTS Montreal, Canada)

Towards Generalizing to Unseen Domains with Few Labels 23691
Chamuditha Jayanga Galappaththige (Queensland University of Technology, Australia), Sanoojan Baliah (Mohamed bin Zayed University of Artificial Intelligence, UAE), Malitha Gunawardhana (University of Auckland, New Zealand), and Muhammad Haris Khan (Mohamed bin Zayed University of Artificial Intelligence, UAE)

Improved Self-Training for Test-Time Adaptation 23701
Jing Ma (Huazhong University of Science and Technology, China)

Source-Free Domain Adaptation with Frozen Multimodal Foundation Model 23711
Song Tang (University of Shanghai for Science and Technology, China), Wenxin Su (University of Shanghai for Science and Technology, China), Mao Ye (University of Electronic Science and Technology of China, China), and Xiatian Zhu (University of Surrey, UK)

Deep Imbalanced Regression via Hierarchical Classification Adjustment 23721
Haipeng Xiong (National University of Singapore) and Angela Yao (National University of Singapore)

A Versatile Framework for Continual Test-Time Domain Adaptation: Balancing Discriminability and Generalizability 23731
Xu Yang (Xidian University), Xuan Chen (Xidian University), Moqi Li (Xidian University), Kun Wei (Xidian University), and Cheng Deng (Xidian University)

DYSON: Dynamic Feature Space Self-Organization for Online Task-Free Class Incremental Learning 23741
Yuhang He (Xi'an Jiaotong University), Yingjie Chen (Xi'an Jiaotong University), Yuhan Jin (Xi'an Jiaotong University), Songlin Dong (Xi'an Jiaotong University), Xing Wei (Xi'an Jiaotong University), and Yihong Gong (Xi'an Jiaotong University)

Test-Time Linear Out-of-Distribution Detection 23752
Ke Fan (Fudan University), Tong Liu (BOE Technology Group), Xingyu Qiu (Fudan University), Yikai Wang (Fudan University), Lian Huai (BOE Technology Group), Zeyu Shangguan (BOE Technology Group), Shuang Gou (BOE Technology Group), Fengjian Liu (BOE Technology Group), Yuqian Fu (Fudan University), Yanwei Fu (Fudan University), and Xingqun Jiang (BOE Technology Group)

APSeg: Auto-Prompt Network for Cross-Domain Few-Shot Semantic Segmentation 23762
Weizhao He (Shenzhen University), Yang Zhang (Shenzhen University), Wei Zhuo (Shenzhen University), Linlin Shen (Shenzhen University), Jiaqi Yang (Shenzhen University; University of Nottingham), Songhe Deng (Shenzhen University), and Liang Sun (Shenzhen University)

LP++: A Surprisingly Strong Linear Probe for Few-Shot CLIP 23773
Yunshi Huang (ÉTS Montréal), Fereshteh Shakeri (ÉTS Montréal), Jose Dolz (ÉTS Montréal), Malik Boudiaf (ÉTS Montréal), Houda Bahig (Université de Montréal), and Ismail Ben Ayed (ÉTS Montréal)

On the Test-Time Zero-Shot Generalization of Vision-Language Models: Do We Really Need Prompt Learning? 23783
Maxime Zanella (UCLouvain, Belgium) and Ismail Ben Ayed (ETS Montreal, Canada)

Discriminative Sample-Guided and Parameter-Efficient Feature Space Adaptation for Cross-Domain Few-Shot Learning 23794
Rashindrie Perera (University of Melbourne, Australia, Peter MacCallum Cancer Centre, Australia) and Saman Halgamuge (University of Melbourne, Australia)

Regularized Parameter Uncertainty for Improving Generalization in Reinforcement Learning 23805
Pehuen Moure (ETH Zurich and University of Zurich, Switzerland), Longbiao Cheng (ETH Zurich and University of Zurich, Switzerland), Joachim Ott (ETH Zurich and University of Zurich, Switzerland), Zuowen Wang (ETH Zurich and University of Zurich, Switzerland), and Shih-Chii Liu (ETH Zurich and University of Zurich, Switzerland)

An Empirical Study of the Generalization Ability of Lidar 3D Object Detectors to Unseen Domains 23815
George Eskandar (University of Stuttgart)

MMA: Multi-Modal Adapter for Vision-Language Models 23826
Lingxiao Yang (Sun Yat-sen University, China), Ru-Yuan Zhang (Shanghai Jiao Tong University, China), Yanchen Wang (Stanford University, USA), and Xiaohua Xie (Sun Yat-sen University, China)

PerAda: Parameter-Efficient Federated Learning Personalization with Generalization Guarantees 23838
Chulin Xie (University of Illinois Urbana-Champaign), De-An Huang (NVIDIA), Wenda Chu (California Institute of Technology), Daguang Xu (NVIDIA), Chaowei Xiao (University of Wisconsin-Madison), Bo Li (University of Chicago), and Anima Anandkumar (California Institute of Technology)

Bayesian Exploration of Pre-trained Models for Low-shot Image Classification 23849
Yibo Miao (Shanghai Jiaotong University, China), Yu Lei (Shanghai Jiaotong University, China), Feng Zhou (Renmin University of China, China), and Zhijie Deng (Shanghai Jiaotong University, China)

NAYER: Noisy Layer Data Generation for Efficient and Effective Data-free Knowledge Distillation 23860
Minh-Tuan Tran (Monash University, Australia), Trung Le (Monash University, Australia), Xuan-May Le (University of Melbourne, Australia), Mehrtash Harandi (Monash University, Australia), Quan Hung Tran (ServiceNow, USA), and Dinh Phung (Monash University, Australia)

Text-Enhanced Data-free Approach for Federated Class-Incremental Learning 23870
Minh-Tuan Tran (Monash University, Australia), Trung Le (Monash University, Australia), Xuan-May Le (University of Melbourne, Australia), Mehrtash Harandi (Monash University, Australia), and Dinh Phung (Monash University, Australia)

Pre-trained Vision and Language Transformers Are Few-Shot Incremental Learners 23881
Keon-Hee Park (Kyung Hee University, Republic of Korea), Kyungwoo Song (Yonsei University, Republic of Korea), and Gyeong-Moon Park (Kyung Hee University, Republic of Korea)

CDMAD: Class-Distribution-Mismatch-Aware Debiasing for Class-Imbalanced Semi-Supervised Learning 23891
Hyuck Lee (KAIST, Korea) and Heeyoung Kim (KAIST, Korea)

TEA: Test-time Energy Adaptation 23901
Yige Yuan (University of Chinese Academy of Sciences, China), Bingbing Xu (Institute of Computing Technology, Chinese Academy of Sciences, China), Liang Hou (Kuaishou Technology), Fei Sun (Institute of Computing Technology, Chinese Academy of Sciences, China), Huawei Shen (Institute of Computing Technology, Chinese Academy of Sciences, China), and Xueqi Cheng (Institute of Computing Technology, Chinese Academy of Sciences, China)

Universal Semi-Supervised Domain Adaptation by Mitigating Common-Class Bias 23912
*Wenyu Zhang (Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A*STAR), Singapore), Qingmu Liu (National University of Singapore (NUS), Singapore), Felix Ong Wei Cong (National University of Singapore (NUS), Singapore), Mohamed Ragab (Institute for Infocomm Research (I2R), Centre for Frontier AI Research (CFAR), Agency for Science, Technology and Research (A*STAR), Singapore), and Chuan-Sheng Foo (Institute for Infocomm Research (I2R), Centre for Frontier AI Research (CFAR), Agency for Science, Technology and Research (A*STAR), Singapore)*

Leveraging Vision-Language Models for Improving Domain Generalization in Image Classification 23922
Sravanti Addepalli (Indian Institute of Science), Ashish Ramayee Asokan (Indian Institute of Science), Lakshay Sharma (Indian Institute of Science), and R. Venkatesh Babu (Indian Institute of Science)

Learning Equi-angular Representations for Online Continual Learning 23933
Minhyuk Seo (Yonsei University, Republic of Korea), Hyunseo Koh (Yonsei University, Republic of Korea), Wonje Jeung (Yonsei University, Republic of Korea), Minjae Lee (Yonsei University, Republic of Korea), San Kim (Yonsei University, Republic of Korea), Hankook Lee (Sungkyunkwan University, Republic of Korea), Sungjun Cho (LG AI Research, Republic of Korea), Sungik Choi (LG AI Research, Republic of Korea), Hyunwoo Kim (Zhejiang Lab, China), and Jonghyun Choi (Seoul National University, Republic of Korea)

Open-Set Domain Adaptation for Semantic Segmentation 23943
Seun-An Choe (Kyung Hee University, Yongin, Republic of Korea), Ah-Hyung Shin (Kyung Hee University, Yongin, Republic of Korea), Keon-Hee Park (Kyung Hee University, Yongin, Republic of Korea), Jinwoo Choi (Kyung Hee University, Yongin, Republic of Korea), and Gyeong-Moon Park (Kyung Hee University, Yongin, Republic of Korea)

Task-Adaptive Saliency Guidance for Exemplar-free Class Incremental Learning 23954
Xialei Liu (Nankai University, China), Jiang-Tian Zhai (Nankai University, China), Andrew D. Bagdanov (University of Florence, Italy), Ke Li (Tencent, China), and Ming-Ming Cheng (Nankai University, China)

Progressive Semantic-Guided Vision Transformer for Zero-Shot Learning 23964
Shiming Chen (Mohamed bin Zayed University of AI), Wenjin Hou (Huazhong University of Science and Technology), Salman Khan (Mohamed bin Zayed University of AI; Australian National University), and Fahad Shahbaz Khan (Mohamed bin Zayed University of AI; Linkoping University)

Unified Entropy Optimization for Open-Set Test-Time Adaptation 23975
Zhengqing Gao (Institute of Automation, Chinese Academy of Sciences, China), Xu-Yao Zhang (Institute of Automation, Chinese Academy of Sciences, China), and Cheng-Lin Liu (Institute of Automation, Chinese Academy of Sciences, China)

FedSelect: Personalized Federated Learning with Customized Selection of Parameters for Fine-Tuning 23985
Rishub Tamirisa (Lapis Labs, USA), Chulin Xie (Lapis Labs, USA; University of Illinois Urbana-Champaign, USA), Wenxuan Bao (Lapis Labs; University of Illinois Urbana-Champaign), Andy Zhou (Lapis Labs, USA), Ron Arel (Lapis Labs, USA), and Aviv Shamsian (Bar-Ilan University)

Dual-Enhanced Coreset Selection with Class-wise Collaboration for Online Blurry Class Incremental Learning 23995
Yutian Luo (Renmin University of China, China), Shiqi Zhao (China Unicom Research Institute, China), Haoran Wu (China Unicom Research Institute, China), and Zhiwu Lu (Renmin University of China, China)

Troika: Multi-Path Cross-Modal Traction for Compositional Zero-Shot Learning 24005
Siteng Huang (Zhejiang University; Westlake University), Biao Gong (Alibaba Group), Yutong Feng (Alibaba Group), Min Zhang (Westlake University), Yiliang Lv (Alibaba Group), and Donglin Wang (Westlake University)

Unveiling the Unknown: Unleashing the Power of Unknown to Known in Open-Set Source-Free Domain Adaptation 24015
Fuli Wan (Xidian University, China), Han Zhao (Xidian University, China), Xu Yang (Xidian University, China), and Cheng Deng (Xidian University, China)

Dual-Consistency Model Inversion for Non-Exemplar Class Incremental Learning 24025
Zihuan Qiu (University of Electronic Science and Technology of China), Yi Xu (Dalian University of Technology), Fanman Meng (University of Electronic Science and Technology of China), Hongliang Li (University of Electronic Science and Technology of China), Linfeng Xu (University of Electronic Science and Technology of China), and Qingbo Wu (University of Electronic Science and Technology of China)

Domain-Rectifying Adapter for Cross-Domain Few-Shot Segmentation 24036
Jiapeng Su (Harbin Institute of Technology, Shenzhen), Qi Fan (Nanjing University), Wenjie Pei (Harbin Institute of Technology, Shenzhen), Guangming Lu (Harbin Institute of Technology, Shenzhen), and Fanglin Chen (Harbin Institute of Technology, Shenzhen)

Overcoming Generic Knowledge Loss with Selective Parameter Update 24046
Wenxuan Zhang (KAUST, Saudi Arabia), Paul Janson (Concordia University, Canada), Rahaf Aljundi (Toyota Motor Europe, Belgium), and Mohamed Elhoseiny (KAUST, Saudi Arabia)

BrainWash: A Poisoning Attack to Forget in Continual Learning 24057
Ali Abbasi (Vanderbilt University), Parsa Nooralinejad (University of California Davis), Hamed Pirsiavash (University of California Davis), and Soheil Kolouri (Vanderbilt University)

Enhancing Visual Continual Learning with Language-Guided Supervision 24068
Bolin Ni (Institute of Automation, Chinese Academy of Sciences, China), Hongbo Zhao (Institute of Automation, Chinese Academy of Sciences, China), Chenghao Zhang (Institute of Automation, Chinese Academy of Sciences, China), Ke Hu (Institute of Automation, Chinese Academy of Sciences, China), Gaofeng Meng (Institute of Automation, Chinese Academy of Sciences, China), Zhaoxiang Zhang (Institute of Automation, Chinese Academy of Sciences, China), and Shiming Xiang (Institute of Automation, Chinese Academy of Sciences, China)

Orals 6A Low-level Vision and Remote Sensing

LDP: Language-driven Dual-Pixel Image Defocus Deblurring Network 24078
Hao Yang (Beijing Institute of Technology, China), Liyuan Pan (Beijing Institute of Technology, China), Yan Yang (Australian National University, Australia), Richard Hartley (Australian National University, Australia), and Miaomiao Liu (Australian National University, Australia)

S²MVTC: a Simple yet Efficient Scalable Multi-View Tensor Clustering 24088
Zhen Long (University of Electronic Science & Technology of China), Qiyuan Wang (University of Electronic Science & Technology of China), Yazhou Ren (University of Electronic Science & Technology of China), Yipeng Liu (University of Electronic Science & Technology of China), and Ce Zhu (University of Electronic Science & Technology of China)

Task-Driven Wavelets using Constrained Empirical Risk Minimization 24098
Eric Marcus (Netherlands Cancer Institute), Ray Sheombarsing (Netherlands Cancer Institute), Jan-Jakob Sonke (Netherlands Cancer Institute), and Jonas Teuwen (Netherlands Cancer Institute)

Image Processing GNN: Breaking Rigidity in Super-Resolution 24108
Yuchuan Tian (Peking University), Hanting Chen (Huawei Noah's Ark Lab), Chao Xu (Peking University), and Yunhe Wang (Huawei Noah's Ark Lab)

DART: Implicit Doppler Tomography for Radar Novel View Synthesis 24118
Tianshu Huang (Carnegie Mellon University, USA), John Miller (Carnegie Mellon University, USA), Akarsh Prabhakara (Carnegie Mellon University, USA), Tao Jin (Carnegie Mellon University, USA), Tarana Laroia (Carnegie Mellon University, USA), Zico Kolter (Carnegie Mellon University and Bosch Research, USA), and Anthony Rowe (Carnegie Mellon University and Bosch Research, USA)

Orals 6B Image & Video Synthesis

Alchemist: Parametric Control of Material Properties with Diffusion Models 24130
Prafull Sharma (MIT), Varun Jampani (Stability AI), Yuanzhen Li (Google Research), Xuhui Jia (Google Research), Dmitry Lagun (Google Research), Fredo Durand (MIT), Bill Freeman (MIT, Google Research), and Mark Matthews (Google Research)

Generative Image Dynamics 24142
Zhengqi Li (Google Research), Richard Tucker (Google Research), Noah Snavely (Google research), and Aleksander Holynski (Google research)

Visual Anagrams: Generating Multi-View Optical Illusions with Diffusion Models 24154
Daniel Geng (University of Michigan, USA), Inbum Park (University of Michigan, USA), and Andrew Owens (University of Michigan, USA)

MonoHair: High-Fidelity Hair Modeling from a Monocular Video 24164
Keyu Wu (Zhejiang University), Lingchen Yang (ETH Zurich), Zhiyi Kuang (Zhejiang University), Yao Feng (ETH Zurich; Max-Planck Institute), Xutao Han (Zhejiang University), Yuefan Shen (Zhejiang University), Hongbo Fu (City University of Hong Kong; Hong Kong University of Science and Technology), Kun Zhou (Zhejiang University), and Youyi Zheng (Zhejiang University)

Analyzing and Improving the Training Dynamics of Diffusion Models 24174
Tero Karras (NVIDIA), Miika Aittala (NVIDIA), Jaakko Lehtinen (NVIDIA; Aalto University), Janne Hellsten (NVIDIA), Timo Aila (NVIDIA), and Samuli Laine (NVIDIA)

Orals 6C Multi-modal Learning

InternVL: Scaling up Vision Foundation Models and Aligning for Generic Visual-Linguistic Tasks 24185
Zhe Chen (Nanjing University, China), Jiannan Wu (The University of Hong Kong, China), Wenhai Wang (The Chinese University of Hong Kong, China), Weijie Su (University of Science and Technology of China, China), Guo Chen (Nanjing University, China), Sen Xing (Tsinghua University, China), Muyan Zhong (Tsinghua University, China), Qinglong Zhang (Shanghai AI Laboratory, China), Xizhou Zhu (Tsinghua University, China), Lewei Lu (SenseTime Research, China), Bin Li (University of Science and Technology of China, China), Ping Luo (The University of Hong Kong, China), Tong Lu (Nanjing University, China), Yu Qiao (Shanghai AI Laboratory, China), and Jifeng Dai (Tsinghua University, China)

Describing Differences in Image Sets with Natural Language 24199
Lisa Dunlap (UC Berkeley), Yuhui Zhang (Stanford), Xiaohan Wang (Stanford), Ruiqi Zhong (UC Berkeley), Trevor Darrell (UC Berkeley), Jacob Steinhardt (UC Berkeley), Joseph E. Gonzalez (UC Berkeley), and Serena Yeung-Levy (Stanford)

NoiseCLR: A Contrastive Learning Approach for Unsupervised Discovery of Interpretable Directions in Diffusion Models 24209
Yusuf Dalva (Virginia Tech, USA) and Pinar Yanardag (Virginia Tech, USA)

MetaCloak: Preventing Unauthorized Subject-driven Text-to-image Diffusion-based Synthesis via Meta-learning 24219
Yixin Liu (Lehigh University), Chenrui Fan (Huazhong University of Science and Technology), Yutong Dai (Lehigh University), Xun Chen (Samsung Research America), Pan Zhou (Huazhong University of Science and Technology), and Lichao Sun (Lehigh University)

EGTR: Extracting Graph from Transformer for Scene Graph Generation 24229
Jinbae Im (NAVER Cloud AI, Republic of Korea), JeongYeon Nam (NAVER Cloud AI, Republic of Korea), Nokyung Park (NAVER, Republic of Korea), Hyungmin Lee (NAVER, Republic of Korea), and Seunghyun Park (NAVER Cloud AI, Republic of Korea)

Poster Session 6

BadCLIP: Trigger-Aware Prompt Learning for Backdoor Attacks on CLIP 24239
Jiawang Bai (Tsinghua University, China), Kuofeng Gao (Tsinghua University, China), Shaobo Min (Tencent, China), Shu-Tao Xia (Tsinghua University, China), Zhifeng Li (Tencent, China), and Wei Liu (Tencent, China)

Semantic-Aware Multi-Label Adversarial Attacks 24251
Hassan Mahmood (Northeastern University, USA) and Ehsan Elhamifar (Northeastern University, USA)

Defense without Forgetting: Continual Adversarial Defense with Anisotropic & Isotropic Pseudo Replay 24263
Yuhang Zhou (Harbin Institute of Technology, Shenzhen) and Zhongyun Hua (Harbin Institute of Technology, Shenzhen)

Learning to Transform Dynamically for Better Adversarial Transferability 24273
Rongyi Zhu (University of Rochester, USA), Zeliang Zhang (University of Rochester, USA), Susan Liang (University of Rochester, USA), Zhuo Liu (University of Rochester, USA), and Chenliang Xu (University of Rochester, USA)

Infrared Adversarial Car Stickers 24284
Xiaopei Zhu (Tsinghua University, China), Yuqiu Liu (Beijing Forestry University, China), Zhanhao Hu (UC Berkeley, USA), Jianmin Li (Tsinghua University, China), and Xiaolin Hu (Tsinghua University, China)

Unsegment Anything by Simulating Deformation 24294
Jiahao Lu (National University of Singapore), Xingyi Yang (National University of Singapore), and Xinchao Wang (National University of Singapore)

Efficient Model Stealing Defense with Noise Transition Matrix 24305
Dong-Dong Wu (Southeast University), Chilin Fu (Ant Group), Weichang Wu (Ant Group), Wenwen Xia (Ant Group), Xiaolu Zhang (Ant Group), Jun Zhou (Ant Group), and Min-Ling Zhang (Southeast University)

Fully Exploiting Every Real Sample: SuperPixel Sample Gradient Model Stealing 24316
Yunlong Zhao (Central South University), Xiaoheng Deng (Central South University), Yijing Liu (Zhejiang University), Xinjun Pei (Central South University), Jiazhi Xia (Central South University), and Wei Chen (Zhejiang University)

Hide in Thicket: Generating Imperceptible and Rational Adversarial Perturbations on 3D Point Clouds 24326
Tianrui Lou (Sun Yat-Sen University), Xiaojun Jia (Nanyang Technological University), Jindong Gu (University of Oxford), Li Liu (National University of Defense Technology), Siyuan Liang (National University of Singapore), Bangyan He (Chinese Academy of Sciences), and Xiaochun Cao (Sun Yat-Sen University)

Boosting Adversarial Transferability by Block Shuffle and Rotation 24336
Kunyu Wang (Chinese University of Hong Kong), Xuanran He (Nanyang Technological University), Wenxuan Wang (Chinese University of Hong Kong), and Xiaosen Wang (Huawei Singularity Security Lab)

Robust Overfitting Does Matter: Test-Time Adversarial Purification With FGSM 24347
Linyu Tang (Chongqing University) and Lei Zhang (Chongqing University)

Data Poisoning based Backdoor Attacks to Contrastive Learning 24357
Jinghuai Zhang (University of California, Los Angeles, USA), Hongbin Liu (Duke University, USA), Jinyuan Jia (Penn State University, USA), and Neil Zhenqiang Gong (Duke University, USA)

NAPGuard: Towards Detecting Naturalistic Adversarial Patches 24367
Siyang Wu (Beihang University, China), Jiakai Wang (Zhongguancun Laboratory, China), Jiejie Zhao (Zhongguancun Laboratory, China), Yazhe Wang (Zhongguancun Laboratory, China), and Xianglong Liu (Beihang University, China; Zhongguancun Laboratory, China; Heifei Comprehensive National Science Center, China)

Ensemble Diversity Facilitates Adversarial Transferability 24377
Bowen Tang (University of Electronic Science and Technology of China, China), Zheng Wang (University of Electronic Science and Technology of China, China), Yi Bin (University of Electronic Science and Technology of China, China), Qi Dou (The Chinese University of Hong Kong, China), Yang Yang (University of Electronic Science and Technology of China, China), and Heng Tao Shen (University of Electronic Science and Technology of China, China)

Revamping Federated Learning Security from a Defender's Perspective: A Unified Defense with Homomorphic Encrypted Data Space 24387
K Naveen Kumar (IIT Hyderabad, India), Reshmi Mitra (SEMO, USA), and C Krishna Mohan (IIT Hyderabad, India)

Can Protective Perturbation Safeguard Personal Data from Being Exploited by Stable Diffusion? 24398
Zhengyue Zhao (University of Chinese Academy of Sciences, China), Jinhao Duan (Drexel University, USA), Kaidi Xu (Drexel University, USA), Chenan Wang (Drexel University, USA), Rui Zhang (Institute of Computing Technology, Chinese Academy of Sciences, China), Zidong Du (Institute of Computing Technology, Chinese Academy of Sciences, China), Qi Guo (Institute of Computing Technology, Chinese Academy of Sciences, China), and Xing Hu (Institute of Computing Technology, Chinese Academy of Sciences, China)

One Prompt Word is Enough to Boost Adversarial Robustness for Pre-trained Vision-Language Models 24408
Lin Li (King's College London, UK), Haoyan Guan (King's College London, UK), Jianing Qiu (Imperial College London, UK), and Michael Spratling (King's College London, UK)

Watermark-embedded Adversarial Examples for Copyright Protection against Diffusion Models 24420
Peifei Zhu (LY Corporation, Japan), Tsubasa Takahashi (LY Corporation, Japan), and Hirokatsu Kataoka (LY Corporation, Japan)

Not All Prompts Are Secure: A Switchable Backdoor Attack Against Pre-trained Vision Transfomers 24431
Sheng Yang (Tsinghua University, China), Jiawang Bai (Tsinghua University, China), Kuofeng Gao (Tsinghua University, China), Yong Yang (Tencent Security Platform Department, China), Yiming Li (Zhejiang University, China), and Shu-Tao Xia (Tsinghua University, China)

Focus on Hiders: Exploring Hidden Threats for Enhancing Adversarial Training 24442
Qian Li (Shanghai Jiao Tong University, Shanghai, China; Ningbo Institute of Digital Twin, Eastern Institute of Technology, Ningbo, China), Yuxiao Hu (The Hong Kong Polytechnic University, HongKong, China; Ningbo Institute of Digital Twin, Eastern Institute of Technology, Ningbo, China), Yinpeng Dong (Institute for AI, Tsinghua-Bosch Joint ML Center, THBI Lab, BNRist Center, Tsinghua University, Beijing, China), Dongxiao Zhang (Ningbo Institute of Digital Twin, Eastern Institute of Technology, Ningbo, China), and Yuntian Chen (Ningbo Institute of Digital Twin, Eastern Institute of Technology, Ningbo, China)

Physical 3D Adversarial Attacks against Monocular Depth Estimation in Autonomous Driving . 24452
Junhao Zheng (Xi'an Jiaotong University, China), Chenhao Lin (Xi'an Jiaotong University, China), Jiahao Sun (Xi'an Jiaotong University, China), Zhengyu Zhao (Xi'an Jiaotong University, China), Qian Li (Xi'an Jiaotong University, China), and Chao Shen (Xi'an Jiaotong University, China)

Distraction is All You Need: Memory-Efficient Image Immunization against Diffusion-Based Image Editing 24462
Ling Lo (National Yang Ming Chiao Tung University), Cheng Yu Yeo (National Yang Ming Chiao Tung University), Hong-Han Shuai (National Yang Ming Chiao Tung University), and Wen-Huang Cheng (National Taiwan University)

PAD: Patch-Agnostic Defense against Adversarial Patch Attacks 24472
Lihua Jing (CAS Institute of Information Engineering, China; University of Chinese Academy of Sciences, China), Rui Wang (CAS Institute of Information Engineering, China; University of Chinese Academy of Sciences, China), Wenqi Ren (Shenzhen Campus of Sun Yat-sen University, China), Xin Dong (CAS Institute of Information Engineering, China; University of Chinese Academy of Sciences, China), and Cong Zou (CAS Institute of Information Engineering, China; University of Chinese Academy of Sciences, China)

PeerAiD: Improving Adversarial Distillation from a Specialized Peer Tutor 24482
Jaewon Jung (Seoul National University, South Korea), Hongsun Jang (Seoul National University, South Korea), Jaeyong Song (Seoul National University, South Korea), and Jinho Lee (Seoul National University, South Korea)

Revisiting Adversarial Training Under Long-Tailed Distributions 24492
Xinli Yue (Wuhan University), Ningping Mou (Wuhan University), Qian Wang (Wuhan University), and Lingchen Zhao (Wuhan University)

Pre-trained Model Guided Fine-Tuning for Zero-Shot Adversarial Robustness 24502
Sibo Wang (Institute of Computing Technology, Chinese Academy of Sciences, China), Jie Zhang (Institute of Computing Technology, Chinese Academy of Sciences, China), Zheng Yuan (Institute of Computing Technology, Chinese Academy of Sciences, China), and Shiguang Shan (Institute of Computing Technology, Chinese Academy of Sciences, China)

Towards Transferable Targeted 3D Adversarial Attack in the Physical World 24512
Yao Huang (Beihang University, China), Yinpeng Dong (Tsinghua University, China), Shouwei Ruan (Beihang University, China), Xiao Yang (Tsinghua University, China), Hang Su (Tsinghua University, China), and Xingxing Wei (Beihang University, China)

Nearest is Not Dearest: Towards Practical Defense against Quantization-conditioned Backdoor Attacks 24523
Boheng Li (Zhejiang University & Wuhan University, China), Yishuo Cai (Central South University, China), Haowei Li (Wuhan University), Feng Xue (X Digital Dynamics, China), Zhifeng Li (Tencent Data Platform, China), and Yiming Li (Zhejiang University, China)

Perturbing Attention Gives You More Bang for the Buck: Subtle Imaging Perturbations That Efficiently Fool Customized Diffusion Models 24534
Jingyao Xu (Beijing Jiaotong University, China), Yuetong Lu (Beijing Jiaotong University, China), Yandong Li (Google Research, USA), Siyang Lu (Beijing Jiaotong University, China), Dongdong Wang (University of Central Florida, USA), and Xiang Wei (Beijing Jiaotong University, China)

Boosting Adversarial Training via Fisher-Rao Norm-based Regularization 24544
Xiangyu Yin (University of Liverpool, UK) and Wenjie Ruan (University of Liverpool, UK)

Random Entangled Tokens for Adversarially Robust Vision Transformer 24554
Huihui Gong (The University of Sydney, Australia; CSIRO Data61, Australia), Minjing Dong (City University of Hong Kong, Hong Kong), Siqi Ma (The University of New South Wales, Australia), Seyit Camtepe (CSIRO Data61, Australia), Surya Nepal (CSIRO Data61, Australia), and Chang Xu (The University of Sydney, Australia)

Backdoor Defense via Test-Time Detecting and Repairing 24564
Jiyang Guan (Institute of Automation, Chinese Academy of Sciences, China), Jian Liang (Institute of Automation, Chinese Academy of Sciences, China), and Ran He (Institute of Automation, Chinese Academy of Sciences, China)

1-Lipschitz Layers Compared: Memory, Speed, and Certifiable Robustness 24574
Bernd Prach (Institute of Science and Technology Austria), Fabio Brau (Scuola Superiore Sant'Anna), Giorgio Buttazzo (Scuola Superiore Sant'Anna), and Christoph H. Lampert (Institute of Science and Technology Austria)

DiffAM: Diffusion-based Adversarial Makeup Transfer for Facial Privacy Protection 24584
Yuhao Sun (University of Science and Technology of China), Lingyun Yu (University of Science and Technology of China), Hongtao Xie (University of Science and Technology of China), Jiaming Li (University of Science and Technology of China), and Yongdong Zhang (University of Science and Technology of China)

DAP: A Dynamic Adversarial Patch for Evading Person Detectors 24595
Amira Guesmi (New York University Abu Dhabi, UAE), Ruitian Ding (NYU Tandon School of Engineering, USA), Muhammad Abdullah Hanif (New York University Abu Dhabi, UAE), Ihsen Alouani (Queen's University Belfast, UK), and Muhammad Shafique (New York University Abu Dhabi, UAE)

Adversarial Distillation Based on Slack Matching and Attribution Region Alignment 24605
Shenglin Yin (Peking University, Chain), Zhen Xiao (Peking University, Chain), Mingxuan Song (Peking University, Chain), and Jieyi Long (Theta Labs, USA)

Improving Transferable Targeted Adversarial Attacks with Model Self-Enhancement 24615
Han Wu (Sun Yat-sen University, China), Guanyan Ou (Sun Yat-sen University, China), Weibin Wu (Sun Yat-sen University, China), and Zibin Zheng (Sun Yat-sen University, China)

On the Robustness of Large Multimodal Models Against Image Adversarial Attacks 24625
Xuanming Cui (Unversity of Central Florida, USA), Alejandro Aparcedo (Unversity of Central Florida, USA), Young Kyun Jang (Meta, USA), and Ser-Nam Lim (Unversity of Central Florida, USA)

Intriguing Properties of Diffusion Models: An Empirical Study of the Natural Attack Capability in Text-to-Image Generative Models 24635
Takami Sato (University of California, Irvine), Justin Yue (University of California, Irvine), Nanze Chen (University of Cambridge), Ningfei Wang (University of California, Irvine), and Qi Alfred Chen (University of California, Irvine)

BadCLIP: Dual-Embedding Guided Backdoor Attack on Multimodal Contrastive Learning 24645
Siyuan Liang (National University of Singapore), Mingli Zhu (The Chinese University of Hong Kong, Shenzhen), Aishan Liu (Beihang University), Baoyuan Wu (The Chinese University of Hong Kong, Shenzhen), Xiaochun Cao (Sun Yat-sen University), and Ee-Chien Chang (National University of Singapore)

MMCert: Provable Defense against Adversarial Attacks to Multi-modal Models 24655
Yanting Wang (The Pennsylvania State University, USA), Hongye Fu (Zhejiang University, China), Wei Zou (The Pennsylvania State University, USA), and Jinyuan Jia (The Pennsylvania State University, USA)

MimicDiffusion: Purifying Adversarial Perturbation via Mimicking Clean Diffusion Model 24665
Kaiyu Song (Sun Yat-sen University), Hanjiang Lai (Sun Yat-sen University), Yan Pan (Sun Yat-sen University), and Jian Yin (Sun Yat-sen University)

Revisiting Adversarial Training at Scale 24675
Zeyu Wang (UC Santa Cruz), Xianhang Li (UC Santa Cruz), Hongru Zhu (UC Santa Cruz), and Cihang Xie (UC Santa Cruz)

Language-Driven Anchors for Zero-Shot Adversarial Robustness 24686
Xiao Li (Tsinghua University, China), Wei Zhang (Tsinghua University, China), Yining Liu (Harbin Institute of Technolog, Weihai, China), Zhanhao Hu (Tsinghua University, China), Bo Zhang (Tsinghua University, China), and Xiaolin Hu (Tsinghua University, China)

Transferable Structural Sparse Adversarial Attack Via Exact Group Sparsity Training 24696
Di Ming (Chongqing University of Technology, China), Peng Ren (Chongqing University of Technology, China), Yunlong Wang (Chongqing University of Technology, China), and Xin Feng (Chongqing University of Technology, China)

Fooling Polarization-Based Vision using Locally Controllable Polarizing Projection 24706
Zhuoxiao Li (The University of Tokyo), Zhihang Zhong (Shanghai Artificial Intelligence Laboratory), Shohei Nobuhara (Kyoto University), Ko Nishino (Kyoto University), and Yinqiang Zheng (The University of Tokyo)

Overload: Latency Attacks on Object Detection for Edge Devices 24716
Erh-Chung Chen (National Tsing Hua University), Pin-Yu Chen (IBM Research), I-Hsin Chung (IBM Research), and Che-Rung Lee (National Tsing Hua University)

Attack To Defend: Exploiting Adversarial Attacks for Detecting Poisoned Models 24726
Samar Fares (Mohamed Bin Zayed University of Artificial Intelligence (MBZUAI)) and Karthik Nandakumar (Mohamed Bin Zayed University of Artificial Intelligence (MBZUAI))

Towards Understanding and Improving Adversarial Robustness of Vision Transformers 24736
Samyak Jain (Indian Institute of Technology (BHU) Varanasi) and Tanima Dutta (Indian Institute of Technology (BHU) Varanasi)

Towards Fairness-Aware Adversarial Learning 24746
Yanghao Zhang (University of Liverpool, UK), Tianle Zhang (University of Liverpool, UK), Ronghui Mu (University of Liverpool, UK), Xiaowei Huang (University of Liverpool, UK), and Wenjie Ruan (University of Liverpool, UK)

Byzantine-robust Decentralized Federated Learning via Dual-domain Clustering and Trust Bootstrapping 24756
Peng Sun (Hunan University, China), Xinyang Liu (The Hong Kong Polytechnic University, China), Zhibo Wang (Zhejiang University, China), and Bo Liu (Shenzhen Institute of Artificial Intelligence and Robotics for Society (AIRS), China)

Towards General Robustness Verification of MaxPool-based Convolutional Neural Networks via Tightening Linear Approximation 24766
Yuan Xiao (Nanjing University, China), Shiqing Ma (University of Massachusetts Amherst, USA), Juan Zhai (University of Massachusetts Amherst, USA), Chunrong Fang (Nanjing University, China), Jinyuan Jia (Pennsylvania State University, USA), and Zhenyu Chen (Nanjing University, China)

Soften to Defend: Towards Adversarial Robustness via Self-Guided Label Refinement 24776
Zhuorong Li (Hangzhou City University, China), Daiwei Yu (Hangzhou City University, China), Lina Wei (Hangzhou City University, China), Canghong Jin (Hangzhou City University, China), Yun Zhang (Hangzhou City University, China), and Sixian Chan (Zhejiang University of Technology)

SlowFormer: Adversarial Attack on Compute and Energy Consumption of Efficient Vision Transformers 24786
K L Navaneet (University of California, Davis), Soroush Abbasi Koohpayegani (University of California, Davis), Essam Sleiman (University of California, Davis; Harvard University), and Hamed Pirsiavash (University of California, Davis)

LOTUS: Evasive and Resilient Backdoor Attacks through Sub-Partitioning 24798
Siyuan Cheng (Purdue University, USA), Guanhong Tao (Purdue University, USA), Yingqi Liu (Microsoft, USA), Guangyu Shen (Purdue University, USA), Shengwei An (Purdue University, USA), Shiwei Feng (Purdue University, USA), Xiangzhe Xu (Purdue University, USA), Kaiyuan Zhang (Purdue University, USA), Shiqing Ma (University of Massachusetts at Amherst, USA), and Xiangyu Zhang (Purdue University, USA)

Deep-TROJ: An Inference Stage Trojan Insertion Algorithm through Efficient Weight Replacement Attack 24810
Sabbir Ahmed (Binghamton University (SUNY), USA), Ranyang Zhou (New Jersey Institute of Technology, USA), Shaahin Angizi (New Jersey Institute of Technology, USA), and Adnan Siraj Rakin (Binghamton University (SUNY), USA)

Semantic Shield: Defending Vision-Language Models Against Backdooring and Poisoning via Fine-grained Knowledge Alignment 24820
Alvi Md Ishmam (Virginia Tech) and Christopher Thomas (Virginia Tech)

Initialization Matters for Adversarial Transfer Learning 24831
Andong Hua (University of California, Santa Barbara, USA), Jindong Gu (University of Oxford, United Kingdom), Zhiyu Xue (University of California, Santa Barbara, USA), Nicholas Carlini (Google, USA), Eric Wong (University of Pennsylvania, USA), and Yao Qin (University of California, Santa Barbara, USA)

Strong Transferable Adversarial Attacks via Ensembled Asymptotically Normal Distribution Learning 24841
Zhengwei Fang (Beijing Jiaotong University), Rui Wang (Beijing Jiaotong University), Tao Huang (Beijing Jiaotong University), and Liping Jing (Beijing Jiaotong University)

HDRFlow: Real-Time HDR Video Reconstruction with Large Motions 24851
Gangwei Xu (Huazhong University of Science and Technology), Yujin Wang (Shanghai AI Laboratory), Jinwei Gu (The Chinese University of Hong Kong), Tianfan Xue (The Chinese University of Hong Kong), and Xin Yang (Huazhong University of Science and Technology)

A Physics-informed Low-rank Deep Neural Network for Blind and Universal Lens Aberration Correction 24861
Jin Gong (Tsinghua University, China), Runzhao Yang (Tsinghua University, China), Weihang Zhang (Tsinghua University, China), Jinli Suo (Tsinghua University, China), and Qionghai Dai (Tsinghua University, China)

Super-Resolution Reconstruction from Bayer-Pattern Spike Streams 24871
Yanchen Dong (Peking University, China), Ruiqin Xiong (Peking University, China), Jian Zhang (Peking University, China), Zhaofei Yu (Peking University, China), Xiaopeng Fan (Harbin Institute of Technology, China), Shuyuan Zhu (University of Electronic Science and Technology of China, China), and Tiejun Huang (Peking University, China)

In2SET: Intra-Inter Similarity Exploiting Transformer for Dual-Camera Compressive Hyperspectral Imaging 24881
Xin Wang (Beijing Institute of Technology), Lizhi Wang (Beijing Institute of Technology), Xiangtian Ma (Beijing Institute of Technology), Maoqing Zhang (Beijing Institute of Technology), Lin Zhu (Beijing Institute of Technology), and Hua Huang (Beijing Normal University)

SuperSVG: Superpixel-based Scalable Vector Graphics Synthesis 24892
Teng Hu (Shanghai Jiao Tong University, China), Ran Yi (Shanghai Jiao Tong University, China), Baihong Qian (Shanghai Jiao Tong University, China), Jiangning Zhang (Youtu Lab, Tencent, China), Paul L. Rosin (Cardiff University, UK), and Yu-Kun Lai (Cardiff University, UK)

Language-driven All-in-one Adverse Weather Removal 24902
Hao Yang (Beijing Institute of Technology, China), Liyuan Pan (Beijing Institute of Technology, China), Yan Yang (Australian National University, Australia), and Wei Liang (Beijing Institute of Technology, China)

Language-guided Image Reflection Separation 24913
Haofeng Zhong (Peking University, China), Yuchen Hong (Peking University, China), Shuchen Weng (Peking University, China), Jinxiu Liang (Peking University, China), and Boxin Shi (Peking University, China)

Time-Efficient Light-Field Acquisition Using Coded Aperture and Events 24923
Shuji Habuchi (Nagoya University, Japan), Keita Takahashi (Nagoya University, Japan), Chihiro Tsutake (Nagoya University, Japan), Toshiaki Fujii (Nagoya University, Japan), and Hajime Nagahara (Osaka University, Japan)

NB-GTR: Narrow-Band Guided Turbulence Removal 24934
Yifei Xia (Peking University, China), Chu Zhou (Peking University, China), Chengxuan Zhu (Peking University, China), Minggui Teng (Peking University, China), Chao Xu (Peking University, China), and Boxin Shi (Peking University, China)

Complementing Event Streams and RGB Frames for Hand Mesh Reconstruction 24944
Jianping Jiang (Peking University, China), Xinyu Zhou (Peking University, China), Bingxuan Wang (Peking University, China), Xiaoming Deng (Chinese Academy of Sciences, China), Chao Xu (Peking University, China), and Boxin Shi (Peking University, China)

Boosting Spike Camera Image Reconstruction from a Perspective of Dealing with Spike Fluctuations 24955
Rui Zhao (Peking University), Ruiqin Xiong (Peking University), Jing Zhao (Peking University), Jian Zhang (Peking University), Xiaopeng Fan (Harbin Institute of Technology), Zhaofei Yu (Peking University), and Tiejun Huang (Peking University)

Frequency-aware Event-based Video Deblurring for Real-World Motion Blur 24966
Taewoo Kim (KAIST), Hoonhee Cho (KAIST), and Kuk-Jin Yoon (KAIST)

Latency Correction for Event-guided Deblurring and Frame Interpolation 24977
Yixin Yang (Peking University, China), Jinxiu Liang (Peking University, China), Bohan Yu (Peking University, China), Yan Chen (SenseTime Research, China), Jimmy S. Ren (SenseTime Research, China), and Boxin Shi (Peking University, China)

Learning to Remove Wrinkled Transparent Film with Polarized Prior 24987
Jiaqi Tang (The Hong Kong University of Science and Technology (Guangzhou)), Ruizheng Wu (SmartMore Corporation), Xiaogang Xu (The Chinese University of Hong Kong), Sixing Hu (SmartMore Corporation), and Ying-Cong Chen (The Hong Kong University of Science and Technology (Guangzhou))

Dispersed Structured Light for Hyperspectral 3D Imaging 24997
Suhyun Shin (POSTECH), Seokjun Choi (POSTECH), Felix Heide (Princeton University), and Seung-Hwan Baek (POSTECH)

Generalized Event Cameras 25007
Varun Sundar (University of Wisconsin-Madison), Matthew Dutson (University of Wisconsin-Madison), Andrei Ardelean (NovoViz), Claudio Bruschini (Ecole Polytechnique Federale de Lausanne), Edoardo Charbon (Ecole Polytechnique Federale de Lausanne), and Mohit Gupta (University of Wisconsin-Madison)

Intensity-Robust Autofocus for Spike Camera 25018
Changqing Su (Peking University), Zhiyuan Ye (Nanchang Hangkong University), Yongsheng Xiao (Nanchang Hangkong University), You Zhou (Nanjing University), Zhen Cheng (Tsinghua University), Bo Xiong (Peking University), Zhaofei Yu (Peking University), and Tiejun Huang (Peking University)

Selective Nonlinearities Removal from Digital Signals 25028
Krzysztof A. Maliszewski (University of Canterbury, New Zealand), Magdalena A. Urbańska (Massey University, New Zealand), Varvara Vetrova (University of Canterbury, New Zealand), and Sylwia M. Kolenderska (University of Canterbury, New Zealand)

Close Imitation of Expert Retouching for Black-and-White Photography 25037
Seunghyun Shin (GIST AI Graduate School, South Korea), Jisu Shin (GIST AI Graduate School, South Korea), Jihwan Bae (Cha University, South Korea), Inwook Shim (Inha University, South Korea), and Hae-Gon Jeon (GIST AI graduate School, South Korea)

Spike-guided Motion Deblurring with Unknown Modal Spatiotemporal Alignment 25047
Jiyuan Zhang (Peking University, China), Shiyan Chen (Peking University, China), Yajing Zheng (Peking University, China), Zhaofei Yu (Peking University, China), and Tiejun Huang (Peking University, China)

Coherence As Texture – Passive Textureless 3D Reconstruction by Self-interference 25058
Wei-Yu Chen (Carnegie Mellon University, USA), Aswin C. Sankaranarayanan (Carnegie Mellon University, USA), Anat Levin (Technion, Israel), and Matthew O'Toole (Carnegie Mellon University, USA)

TurboSL: Dense, Accurate and Fast 3D by Neural Inverse Structured Light 25067
Parsa Mirdehghan (University of Toronto, Canada; Vector Institute, Canada), Maxx Wu (University of Toronto, Canada; Vector Institute, Canada), Wenzheng Chen (University of Toronto, Canada; Vector Institute, Canada), David B. Lindell (University of Toronto, Canada; Vector Institute, Canada), and Kiriakos N. Kutulakos (University of Toronto, Canada; Vector Institute, Canada)

SPIDeRS: Structured Polarization for Invisible Depth and Reflectance Sensing 25077
Tomoki Ichikawa (Kyoto University, Japan), Shohei Nobuhara (Kyoto University, Japan), and Ko Nishino (Kyoto University, Japan)

CPP-Net: Embracing Multi-Scale Feature Fusion into Deep Unfolding CP-PPA Network for Compressive Sensing 25086
Zhen Guo (Northwestern Polytechnical University, China) and Hongping Gan (Northwestern Polytechnical University, China)

SwitchLight: Co-design of Physics-driven Architecture and Pre-training Framework for Human Portrait Relighting 25096
Hoon Kim (Beeble AI), Minje Jang (Beeble AI), Wonjun Yoon (Beeble AI), Jisoo Lee (Beeble AI), Donghyun Na (Beeble AI), and Sanghyun Woo (New York University)

Diffeomorphic Template Registration for Atmospheric Turbulence Mitigation 25107
Dong Lao (UCLA, USA), Congli Wang (UC Berkeley, USA), Alex Wong (Yale, UCLA), and Stefano Soatto (UCLA, USA)

Towards HDR and HFR Video from Rolling-Mixed-Bit Spikings 25117
Yakun Chang (Beijing Jiaotong University), Yeliduosi Xiaokaiti (Peking University), Yujia Liu (Peking University), Bin Fan (Peking University), Zhaojun Huang (Peking University), Tiejun Huang (Peking University), and Boxin Shi (Peking University)

Progressive Divide-and-Conquer via Subsampling Decomposition for Accelerated MRI 25128
Chong Wang (Nanyang Technological University, Singapore), Lanqing Guo (Nanyang Technological University, Singapore), Yufei Wang (Nanyang Technological University, Singapore), Hao Cheng (Nanyang Technological University, Singapore), Yi Yu (Nanyang Technological University, Singapore), and Bihan Wen (Nanyang Technological University, Singapore)

Generative Quanta Color Imaging 25138
Vishal Purohit (Purdue University, USA), Junjie Luo (Purdue University, USA), Yiheng Chi (Purdue University, USA), Qi Guo (Purdue University, USA), Stanley H. Chan (Purdue University, USA), and Qiang Qiu (Purdue University, USA)

UFC-Net: Unrolling Fixed-point Continuous Network for Deep Compressive Sensing 25149
Xiaoyang Wang (Northwestern Polytechnical University, China) and Hongping Gan (Northwestern Polytechnical University, China)

Batch Normalization Alleviates the Spectral Bias in Coordinate Networks 25160
Zhicheng Cai (Nanjing University, China), Hao Zhu (Nanjing University, China), Qiu Shen (Nanjing University, China), Xinran Wang (Nanjing University, China), and Xun Cao (Nanjing University, China)

EVS-assisted Joint Deblurring, Rolling-Shutter Correction and Video Frame Interpolation through Sensor Inverse Modeling 25172
Rui Jiang (OMNIVISION, Singapore), Fangwen Tu (OMNIVISION, Singapore), Yixuan Long (OMNIVISION, Singapore), Aabhaas Vaish (OMNIVISION, USA), Bowen Zhou (OMNIVISION, China), Qinyi Wang (OMNIVISION, Singapore), Wei Zhang (OMNIVISION, China), Yuntan Fang (OMNIVISION, China), Luis Eduardo Garcia Capel (OMNIVISION, Singapore), Bo Mu (OMNIVISION, USA), Tiejun Dai (OMNIVISION, USA), and Andreas Suess (OMNIVISION, USA)

Unsupervised Deep Unrolling Networks for Phase Unwrapping 25182
Zhile Chen (South China University of Technology, China), Yuhui Quan (South China University of Technology, China), and Hui Ji (National University of Singapore, Singapore)

LAN: Learning to Adapt Noise for Image Denoising 25193
Changjin Kim (Hanyang University), Tae Hyun Kim (Hanyang University), and Sungyong Baik (Hanyang University)

Snapshot Lidar: Fourier Embedding of Amplitude and Phase for Single-Image Depth Reconstruction 25203
Sarah Friday (Dartmouth College), Yunzi Shi (Dartmouth College), Yaswanth Cherivirala (University of Michigan), Vishwanath Saragadam (University of California, Riverside), and Adithya Pediredla (Dartmouth College)

FC-GNN: Recovering Reliable and Accurate Correspondences from Interferences 25213
Haobo Xu (Shanghai Jiao Tong University), Jun Zhou (Shanghai Jiao Tong University), Hua Yang (Shanghai Jiao Tong University), Renjie Pan (Shanghai Jiao Tong University), and Cunyan Li (Shanghai Jiao Tong University)

Projecting Trackable Thermal Patterns for Dynamic Computer Vision 25223
Mark Sheinin (Carnegie Mellon University, USA), Aswin C. Sankaranarayanan (Carnegie Mellon University, USA), and Srinivasa G. Narasimhan (Carnegie Mellon University, USA)

PixelRNN: In-pixel Recurrent Neural Networks for End-to-end–optimized Perception with Neural Sensors 25233
Haley M. So (Stanford University, USA), Laurie Bose (The University of Manchester, England), Piotr Dudek (The University of Manchester, England), and Gordon Wetzstein (Stanford University, USA)

Image Restoration by Denoising Diffusion Models with Iteratively Preconditioned Guidance 25245
Tomer Garber (Open University of Israel) and Tom Tirer (Bar-Ilan University)

Equivariant Plug-and-Play Image Reconstruction 25255
Matthieu Terris (Université Paris-Saclay, Inria, CEA), Thomas Moreau (Université Paris-Saclay, Inria, CEA), Nelly Pustelnik (CNRS, ENS Lyon), and Julian Tachella (CNRS, ENS Lyon)

CodedEvents: Optimal Point-Spread-Function Engineering for 3D-Tracking with Event Cameras 25265
Sachin Shah (University of Maryland, College Park), Matthew A. Chan (University of Maryland, College Park), Haoming Cai (University of Maryland, College Park), Jingxi Chen (University of Maryland, College Park), Sakshum Kulshrestha (University of Maryland, College Park), Chahat Deep Singh (University of Maryland, College Park), Yiannis Aloimonos (University of Maryland, College Park), and Christopher A. Metzler (University of Maryland, College Park)

WaveMo: Learning Wavefront Modulations to See Through Scattering 25276
Mingyang Xie (University of Maryland, USA), Haiyun Guo (Rice University, USA), Brandon Y. Feng (Massachusetts Institute of Technology), Lingbo Jin (Rice University, USA), Ashok Veeraraghavan (Rice University, USA), and Christopher A. Metzler (University of Maryland, USA)

Turb-Seg-Res: A Segment-then-Restore Pipeline for Dynamic Videos with Atmospheric Turbulence 25286
Ripon Kumar Saha (Arizona State University, USA), Dehao Qin (Clemson University, USA), Nianyi Li (Clemson University, USA), Jinwei Ye (George Mason University, USA), and Suren Jayasuriya (Arizona State University, USA)

DiffSCI: Zero-Shot Snapshot Compressive Imaging via Iterative Spectral Diffusion Model 25297
Zhenghao Pan (Harbin Institute of Technology (Shenzhen), China), Haijin Zeng (IMEC-UGent, Belgium), Jiezhang Cao (ETH Zurich, Switzerland), Kai Zhang (Nanjing University, China), and Yongyong Chen (Harbin Institute of Technology (Shenzhen), China)

Resolution Limit of Single-Photon LiDAR 25307
Stanley H. Chan (Purdue University, USA), Hashan K. Weerasooriya (Purdue University, USA), Weijian Zhang (Purdue University), Pamela Abshire (University of Maryland, College Park, USA), Istvan Gyongy (University of Edinburgh, UK), and Robert K. Henderson (University of Edinburgh, UK)

QN-Mixer: A Quasi-Newton MLP-Mixer Model for Sparse-View CT Reconstruction 25317
Ishak Ayad (ETIS & AGM, CY Cergy Paris University, ENSEA, CNRS), Nicolas Larue (ETIS , CY Cergy Paris University, ENSEA, CNRS, University of Ljubljana), and Mai K. Nguyen (ETIS , CY Cergy Paris University, ENSEA, CNRS)

Dual-Scale Transformer for Large-Scale Single-Pixel Imaging 25327
Gang Qu (Westlake University, China), Ping Wang (Westlake University; Zhejiang University, China), and Xin Yuan (Westlake University, China)

Rolling Shutter Correction with Intermediate Distortion Flow Estimation 25338
Mingdeng Cao (The University of Tokyo), Sidi Yang (Tsinghua University), Yujiu Yang (Tsinghua University), and Yinqiang Zheng (The University of Tokyo)

Passive Snapshot Coded Aperture Dual-Pixel RGB-D Imaging 25348
Bhargav Ghanekar (Rice University), Salman Siddique Khan (Rice University), Pranav Sharma (IIT Madras), Shreyas Singh (IIT Madras), Vivek Boominathan (Rice University), Kaushik Mitra (IIT Madras), and Ashok Veeraraghavan (Rice University)

Single View Refractive Index Tomography with Neural Fields 25358
Brandon Zhao (Caltech), Aviad Levis (Caltech), Liam Connor (Cahill Center for Astronomy and Astrophysics, Caltech), Pratul P. Srinivasan (Google Research), and Katherine L. Bouman (Caltech)

SPECAT: SPatial-spEctral Cumulative-Attention Transformer for High-Resolution Hyperspectral Image Reconstruction 25368
Zhiyang Yao (Tsinghua University), Shuyang Liu (Tsinghua University), Xiaoyun Yuan (Tsinghua University), and Lu Fang (Tsinghua University)

Fourier Priors-Guided Diffusion for Zero-Shot Joint Low-Light Enhancement and Deblurring ... 25378
Xiaoqian Lv (Harbin Institute of Technology, China), Shengping Zhang (Harbin Institute of Technology, China), Chenyang Wang (Harbin Institute of Technology, China), Yichen Zheng (Huazhong University of Science and Technology, China), Bineng Zhong (Guangxi Normal University, China), Chongyi Li (Nankai University, China), and Liqiang Nie (Harbin Institute of Technology, China)

Color Shift Estimation-and-Correction for Image Enhancement 25389
Yiyu Li (City University of Hong Kong), Ke Xu (City University of Hong Kong), Gerhard Petrus Hancke (City University of Hong Kong), and Rynson W.H. Lau (City University of Hong Kong)

Video Super-Resolution Transformer with Masked Inter&Intra-Frame Attention 25399
Xingyu Zhou (University of Electronic Science and Technology of China), Leheng Zhang (University of Electronic Science and Technology of China), Xiaorui Zhao (University of Electronic Science and Technology of China), Keze Wang (Sun Yat-sen University), Leida Li (Xidian University), and Shuhang Gu (University of Electronic Science and Technology of China)

Distilling Semantic Priors from SAM to Efficient Image Restoration Models 25409
Quan Zhang (Tsinghua Shenzhen International Graduate School, China; Huawei Noah's Ark Lab, China), Xiaoyu Liu (University of Science and Technology of China, China; Huawei Noah's Ark Lab, China), Wei Li (Huawei Noah's Ark Lab, China), Hanting Chen (Huawei Noah's Ark Lab, China), Junchao Liu (Huawei Noah's Ark Lab, China), Jie Hu (Huawei Noah's Ark Lab, China), Zhiwei Xiong (University of Science and Technology of China, China), Chun Yuan (Tsinghua Shenzhen International Graduate School, China), and Yunhe Wang (Huawei Noah's Ark Lab, China)

Beyond Average: Individualized Visual Scanpath Prediction 25420
Xianyu Chen (University of Minnesota, USA), Ming Jiang (University of Minnesota, USA), and Qi Zhao (University of Minnesota, USA)

Multimodal Prompt Perceiver: Empower Adaptiveness, Generalizability and Fidelity for All-in-One Image Restoration 25432
Yuang Ai (Institute of Automation, Chinese Academy of Sciences, China), Huaibo Huang (Institute of Automation, Chinese Academy of Sciences, China), Xiaoqiang Zhou (University of Science and Technology of China, China), Jiexiang Wang (University of Science and Technology of China, China), and Ran He (Institute of Automation, Chinese Academy of Sciences, China)

Selective Hourglass Mapping for Universal Image Restoration Based on Diffusion Model 25445
Dian Zheng (Sun Yat-sen University), Xiao-Ming Wu (Sun Yat-sen University), Shuzhou Yang (Peking University), Jian Zhang (Peking University), Jian-Fang Hu (Sun Yat-sen University), and Wei-Shi Zheng (Sun Yat-sen University)

SeeSR: Towards Semantics-Aware Real-World Image Super-Resolution 25456
Rongyuan Wu (The HongKong Polytechnic University; OPPO Research Institute), Tao Yang (ByteDance Inc.), Lingchen Sun (The HongKong Polytechnic University; OPPO Research Institute), Zhengqiang Zhang (The HongKong Polytechnic University; OPPO Research Institute), Shuai Li (The HongKong Polytechnic University; OPPO Research Institute), and Lei Zhang (The HongKong Polytechnic University; OPPO Research Institute)

Revisiting Single Image Reflection Removal In the Wild 25468
Yurui Zhu (University of Science and Technology of China), Xueyang Fu (University of Science and Technology of China), Peng-Tao Jiang (vivo Mobile Communication Co., Ltd), Hao Zhang (vivo Mobile Communication Co., Ltd), Qibin Sun (University of Science and Technology of China), Jinwei Chen (vivo Mobile Communication Co., Ltd), Zheng-Jun Zha (University of Science and Technology of China), and Bo Li (vivo Mobile Communication Co., Ltd.)

ODCR: Orthogonal Decoupling Contrastive Regularization for Unpaired Image Dehazing 25479
Zhongze Wang (East China University of Science and Technology, China), Haitao Zhao (East China University of Science and Technology, China), Jingchao Peng (East China University of Science and Technology, China), Lujian Yao (East China University of Science and Technology, China), and Kaijie Zhao (East China University of Science and Technology, China)

Q-Instruct: Improving Low-level Visual Abilities for Multi-modality Foundation Models 25490
*Haoning Wu (S-Lab, Nanyang Technological University), Zicheng Zhang (Shanghai Jiaotong University), Erli Zhang (S-Lab, Nanyang Technological University), Chaofeng Chen (S-Lab, Nanyang Technological University), Liang Liao (S-Lab, Nanyang Technological University), Annan Wang (S-Lab, Nanyang Technological University), Kaixin Xu (A*STAR), Chunyi Li (Shanghai Jiaotong University), Jingwen Hou (S-Lab, Nanyang Technological University), Guangtao Zhai (Shanghai Jiaotong University), Geng Xue (A*STAR), Wenxiu Sun (Sensetime Research), Qiong Yan (Sensetime Research), and Weisi Lin (S-Lab, Nanyang Technological University)*

Enhancing Quality of Compressed Images by Mitigating Enhancement Bias Towards Compression Domain 25501
Qunliang Xing (Beihang University, China), Mai Xu (Beihang University, China), Shengxi Li (Beihang University, China), Xin Deng (Beihang University, China), Meisong Zheng (Alibaba Group, China), Huaida Liu (Alibaba Group, China), and Ying Chen (Alibaba Group, China)

Attentive Illumination Decomposition Model for Multi-Illuminant White Balancing 25512
Dongyoung Kim (Yonsei University, Republic of Korea), Jinwoo Kim (Yonsei University, Republic of Korea), Junsang Yu (Samsung Advanced Institute of Technology), and Seon Joo Kim (Yonsei University, Republic of Korea)

NightCC: Nighttime Color Constancy via Adaptive Channel Masking 25522
Shuwei Li (National University of Singapore) and Robby T. Tan (ASUS Intelligent Cloud Services; National University of Singapore)

Navigating Beyond Dropout: An Intriguing Solution towards Generalizable Image Super Resolution 25532
Hongjun Wang (The University of Tokyo), Jiyuan Chen (The Hong Kong Polytechnic University), Yinqiang Zheng (The University of Tokyo), and Tieyong Zeng (The Chinese University of Hong Kong)

Learning Inclusion Matching for Animation Paint Bucket Colorization 25544
Yuekun Dai (S-Lab, Nanyang Technological University), Shangchen Zhou (S-Lab, Nanyang Technological University), Qinyue Li (S-Lab, Nanyang Technological University), Chongyi Li (S-Lab, Nanyang Technological University), and Chen Change Loy (S-Lab, Nanyang Technological University)

Defense Against Adversarial Attacks on No-Reference Image Quality Models with Gradient Norm Regularization 25554
Yujia Liu (Peking University, China), Chenxi Yang (Peking University, China), Dingquan Li (Peng Cheng Laboratory, China), Jianhao Ding (Peking University, China), and Tingting Jiang (Peking University, China)

Towards Backward-Compatible Continual Learning of Image Compression 25564
Zhihao Duan (Purdue University, USA), Ming Lu (Nanjing University, China), Justin Yang (Purdue University, USA), Jiangpeng He (Purdue University, USA), Zhan Ma (Nanjing University, China), and Fengqing Zhu (Purdue University, USA)

APISR: Anime Production Inspired Real-World Anime Super-Resolution 25574
Boyang Wang (University of Michigan, Ann Arbor, USA), Fengyu Yang (Yale University, USA), Xihang Yu (University of Michigan, Ann Arbor, USA), Chao Zhang (Zhejiang University, China), and Hanbin Zhao (Zhejiang University, China)

Unifying Automatic and Interactive Matting with Pretrained ViTs 25585
Zixuan Ye (Huazhong University of Science and Technology), Wenze Liu (Huazhong University of Science and Technology), He Guo (Huazhong University of Science and Technology), Yujia Liang (Huazhong University of Science and Technology), Chaoyi Hong (Huazhong University of Science and Technology), Hao Lu (Huazhong University of Science and Technology), and Zhiguo Cao (Huazhong University of Science and Technology)

Motion-adaptive Separable Collaborative Filters for Blind Motion Deblurring 25595
Chengxu Liu (Xi'an Jiaotong University, China, Shaanxi Yulan Jiuzhou Intelligent Optoelectronic Technology Co., Ltd, China), Xuan Wang (MEGVII Technology, China), Xiangyu Xu (Xi'an Jiaotong University, China), Ruhao Tian (Xi'an Jiaotong University, China), Shuai Li (MEGVII Technology, China), Xueming Qian (Xi'an Jiaotong University, China, Shaanxi Yulan Jiuzhou Intelligent Optoelectronic Technology Co., Ltd, China), and Ming-Hsuan Yang (University of California, Merced, USA)

Genuine Knowledge from Practice: Diffusion Test-Time Adaptation for Video Adverse Weather Removal 25606
Yijun Yang (The Hong Kong University of Science and Technology (Guangzhou), China), Hongtao Wu (The Hong Kong University of Science and Technology (Guangzhou), China), Angelica I. Aviles-Rivero (University of Cambridge, United Kindom), Yulun Zhang (ETH Zürich, Switzerland), Jing Qin (The Hong Kong Polytechnic University, China), and Lei Zhu (The Hong Kong University of Science and Technology (Guangzhou), China)

HomoFormer: Homogenized Transformer for Image Shadow Removal 25617
Jie Xiao (University of Science and Technology of China, China), Xueyang Fu (University of Science and Technology of China, China), Yurui Zhu (University of Science and Technology of China, China), Dong Li (University of Science and Technology of China, China), Jie Huang (University of Science and Technology of China, China), Kai Zhu (Alibaba Group), and Zheng-Jun Zha (University of Science and Technology of China, China)

Bidirectional Multi-Scale Implicit Neural Representations for Image Deraining 25627
Xiang Chen (Nanjing University of Science and Technology, China), Jinshan Pan (Nanjing University of Science and Technology, China), and Jiangxin Dong (Nanjing University of Science and Technology, China)

LED: A Large-scale Real-world Paired Dataset for Event Camera Denoising 25637
Yuxing Duan (Huazhong University of Science and Technology)

Seeing Motion at Nighttime with an Event Camera 25648
Haoyue Liu (Huazhong University of Science and Technology), Shihan Peng (Huazhong University of Science and Technology), Lin Zhu (Beijing Institute of Technology), Yi Chang (Huazhong University of Science and Technology), Hanyu Zhou (Huazhong University of Science and Technology), and Luxin Yan (Huazhong University of Science and Technology)

Leveraging Frame Affinity for sRGB-to-RAW Video De-rendering 25659
Chen Zhang (SenseTime Research and Tetras.AI, China), Wencheng Han (University of Macau, China), Yang Zhou (SenseTime Research and Tetras.AI, China), Jianbing Shen (University of Macau, China), Cheng-zhong Xu (University of Macau, China), and Wentao Liu (SenseTime Research and Tetras.AI, China)

Scaling Up to Excellence: Practicing Model Scaling for Photo-Realistic Image Restoration In the Wild 25669
Fanghua Yu (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences), Jinjin Gu (Shanghai AI Laboratory), Zheyuan Li (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences), Jinfan Hu (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences), Xiangtao Kong (The Hong Kong Polytechnic University), Xintao Wang (ARC Lab, Tencent PCG), Jingwen He (The Chinese University of Hong Kong; Shanghai AI Laboratory), Yu Qiao (Shanghai AI Laboratory), and Chao Dong (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences; Shanghai AI Laboratory)

AdaRevD: Adaptive Patch Exiting Reversible Decoder Pushes the Limit of Image Deblurring 25681
Xintian Mao (East China Normal University, China), Qingli Li (East China Normal University, China), and Yan Wang (East China Normal University, China)

Unsupervised Blind Image Deblurring Based on Self-Enhancement 25691
Lufei Chen (Sichuan University, China), Xiangpeng Tian (Sichuan University, China), Shuhua Xiong (Sichuan University, China), Yinjie Lei (Sichuan University, China), and Chao Ren (Sichuan University, China)

TTA-EVF: Test-Time Adaptation for Event-based Video Frame Interpolation via Reliable Pixel and Sample Estimation 25701
Hoonhee Cho (KAIST), Taewoo Kim (KAIST), Yuhwan Jeong (KAIST), and Kuk-Jin Yoon (KAIST)

Learning Coupled Dictionaries from Unpaired Data for Image Super-Resolution 25712
Longguang Wang (Aviation University of Air Force), Juncheng Li (Shanghai University), Yingqian Wang (National University of Defense Technology), Qingyong Hu (Independent Researcher), and Yulan Guo (Sun Yat-sen University)

Empowering Resampling Operation for Ultra-High-Definition Image Enhancement with Model-Aware Guidance 25722
Wei Yu (University of Science and Technology of China, China), Jie Huang (University of Science and Technology of China, China), Bing Li (University of Science and Technology of China, China), Kaiwen Zheng (University of Science and Technology of China, China), Qi Zhu (University of Science and Technology of China, China), Man Zhou (University of Science and Technology of China, China), and Feng Zhao (University of Science and Technology of China, China)

Generating Content for HDR Deghosting from Frequency View 25732
Tao Hu (Northwestern Polytechnical University), Qingsen Yan (Northwestern Polytechnical University), Yuankai Qi (Macquarie University), and Yanning Zhang (Northwestern Polytechnical University)

Dual Prior Unfolding for Snapshot Compressive Imaging 25742
Jiancheng Zhang (Northwestern Polytechnical University, China), Haijin Zeng (IMEC & Universiteit Gent, Belgium), Jiezhang Cao (ETH Zürich, Switzerland), Yongyong Chen (Harbin Institute of Technology (Shenzhen), China), Dengxiu Yu (Northwestern Polytechnical University, China), and Yin-Ping Zhao (Northwestern Polytechnical University, China)

Binarized Low-light Raw Video Enhancement 25753
Gengchen Zhang (Beijing Institute of Technology), Yulun Zhang (Shanghai Jiao Tong University), Xin Yuan (Westlake University), and Ying Fu (Beijing Institute of Technology)

Neural Spline Fields for Burst Image Fusion and Layer Separation 25763
Ilya Chugunov (Princeton University), David Shustin (Princeton University), Ruyu Yan (Princeton University), Chenyang Lei (Princeton University), and Felix Heide (Princeton University)

Learning Degradation-Independent Representations for Camera ISP Pipelines 25774
Yanhui Guo (McMaster University, Canada), Fangzhou Luo (McMaster University, Canada), and Xiaolin Wu (McMaster University, Canada)

SeD: Semantic-Aware Discriminator for Image Super-Resolution 25784
Bingchen Li (University of Science and Technology of China), Xin Li (University of Science and Technology of China), Hanxin Zhu (University of Science and Technology of China), Yeying Jin (National University of Singapore), Ruoyu Feng (University of Science and Technology of China), Zhizheng Zhang (Galbot), and Zhibo Chen (University of Science and Technology of China)

SinSR: Diffusion-Based Image Super-Resolution in a Single Step 25796
Yufei Wang (Nanyang Technological University, Singapore), Wenhan Yang (PengCheng Laboratory, China), Xinyuan Chen (Shanghai Artificial Intelligence Laboratory, China), Yaohui Wang (Shanghai Artificial Intelligence Laboratory, China), Lanqing Guo (Nanyang Technological University, Singapore), Lap-Pui Chau (The Hong Kong Polytechnic University, Hong Kong SAR), Ziwei Liu (Nanyang Technological University, Singapore), Yu Qiao (Shanghai Artificial Intelligence Laboratory, China), Alex C. Kot (Nanyang Technological University, Singapore), and Bihan Wen (Nanyang Technological University, Singapore)

Self-Adaptive Reality-Guided Diffusion for Artifact-Free Super-Resolution 25806
Qingping Zheng (Northwestern Polytechnical University), Ling Zheng (Tsinghua-Fuzhou Institute for Data Technology), Yuanfan Guo (Huawei Noah's Ark Lab), Ying Li (Northwestern Polytechnical University), Songcen Xu (Huawei Noah's Ark Lab), Jiankang Deng (Huawei Noah's Ark Lab), and Hang Xu (Huawei Noah's Ark Lab)

Improving Spectral Snapshot Reconstruction with Spectral-Spatial Rectification 25817
Jiancheng Zhang (Northwestern Polytechnical University, China), Haijin Zeng (IMEC & Universiteit Gent, Belgium), Yongyong Chen (Harbin Institute of Technology (Shenzhen), China), Dengxiu Yu (Northwestern Polytechnical University, China), and Yin-Ping Zhao (Northwestern Polytechnical University, China)

Diffusion-based Blind Text Image Super-Resolution 25827
Yuzhe Zhang (Beijing Institute of Technology, China), Jiawei Zhang (SenseTime Research, China), Hao Li (SenseTime Research, China), Zhouxia Wang (The University of Hong Kong, China), Luwei Hou (SenseTime Research, China), Dongqing Zou (SenseTime Research, China), and Liheng Bian (Beijing Institute of Technology, China)

CAMixerSR: Only Details Need More "Attention" 25837
Yan Wang (Nankai-Baidu Joint Lab, Nankai University), Yi Liu (Bytedance Inc.), Shijie Zhao (Bytedance Inc.), Junlin Li (Bytedance Inc.), and Li Zhang (Bytedance Inc.)

ID-Blau: Image Deblurring by Implicit Diffusion-based reBLurring AUgmentation 25847
Jia-Hao Wu (National Yang Ming Chiao Tung University), Fu-Jen Tsai (National Tsing Hua University), Yan-Tsung Peng (National Chengchi University), Chung-Chi Tsai (Qualcomm Technologies, Inc.), Chia-Wen Lin (National Tsing Hua University), and Yen-Yu Lin (National Yang Ming Chiao Tung University)

Low-Res Leads the Way: Improving Generalization for Super-Resolution by Self-Supervised Learning 25857
Haoyu Chen (The Hong Kong University of Science and Technology (Guangzhou), China), Wenbo Li (Huawei, China), Jinjin Gu (The University of Sydney, Australia), Jingjing Ren (The Hong Kong University of Science and Technology (Guangzhou), China), Haoze Sun (Tsinghua University, China), Xueyi Zou (Huawei, China), Zhensong Zhang (Huawei, China), Youliang Yan (Huawei, China), and Lei Zhu (The Hong Kong University of Science and Technology (Guangzhou), China)

CoSeR: Bridging Image and Language for Cognitive Super-Resolution 25868
Haoze Sun (Tsinghua University, China), Wenbo Li (Huawei, China), Jianzhuang Liu (Huawei, China), Haoyu Chen (The Hong Kong University of Science and Technology (Guangzhou), China), Renjing Pei (Huawei, China), Xueyi Zou (Huawei, China), Youliang Yan (Huawei, China), and Yujiu Yang (Tsinghua University, China)

Real-World Efficient Blind Motion Deblurring via Blur Pixel Discretization 25879
Insoo Kim (Samsung Advanced Institute of Technology, South Korea; Korea Advanced Institute of Science and Technology, South Korea), Jae Seok Choi (Samsung Advanced Institute of Technology, South Korea), Geonseok Seo (Samsung Advanced Institute of Technology, South Korea), Kinam Kwon (Samsung Advanced Institute of Technology, South Korea), Jinwoo Shin (Korea Advanced Institute of Science and Technology, South Korea), and Hyong-Euk Lee (Samsung Advanced Institute of Technology, South Korea)

SeNM-VAE: Semi-Supervised Noise Modeling with Hierarchical Variational Autoencoder 25889
Dihan Zheng (Tsinghua University, China), Yihang Zou (Tsinghua University, China), Xiaowen Zhang (Hisilicon, China), and Chenglong Bao (Tsinghua University, China; Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, China)

Text-guided Explorable Image Super-resolution 25900
Kanchana Vaishnavi Gandikota (University of Siegen) and Paramanand Chandramouli (University of Siegen)

Equivariant Multi-Modality Image Fusion 25912
Zixiang Zhao (Xi'an Jiaotong University, China; ETH Zurich, Switzerland), Haowen Bai (Xi'an Jiaotong University, China), Jiangshe Zhang (Xi'an Jiaotong University, China), Yulun Zhang (Shanghai Jiao Tong University, China), Kai Zhang (Nanjing University, China), Shuang Xu (Northwestern Polytechnical University, China), Dongdong Chen (Heriot-Watt University, UK), Radu Timofte (ETH Zurich, Switzerland; University of Wurzburg, Germany), and Luc Van Gool (ETH Zurich, Switzerland; INSAIT, Bulgaria)

Revisiting Spatial-Frequency Information Integration from a Hierarchical Perspective for Panchromatic and Multi-Spectral Image Fusion 25922
Jiangtong Tan (University of Science and Technology of China, China), Jie Huang (University of Science and Technology of China, China), Naishan Zheng (University of Science and Technology of China, China), Man Zhou (University of Science and Technology of China, China), Keyu Yan (University of Science and Technology of China, China), Danfeng Hong (Chinese Academy of Sciences, China), and Feng Zhao (University of Science and Technology of China, China)

MCNet: Rethinking the Core Ingredients for Accurate and Efficient Homography Estimation 25932
Haokai Zhu (Zhejiang University, China), Si-Yuan Cao (Zhejiang University, China), Jianxin Hu (Zhejiang University, China), Sitong Zuo (Beijing University of Posts and Telecommunications, China), Beinan Yu (Zhejiang University, China), Jiacheng Ying (Zhejiang University, China), Junwei Li (Zhejiang University, China), and Hui-Liang Shen (Zhejiang University, China)

Contrastive Pre-Training with Multi-View Fusion for No-Reference Point Cloud Quality Assessment 25942
Ziyu Shan (Shanghai Jiao Tong University, China), Yujie Zhang (Shanghai Jiao Tong University, China), Qi Yang (Tencent, China), Haichen Yang (Shanghai Jiao Tong University), Yiling Xu (Shanghai Jiao Tong University), Jenq-Neng Hwang (University of Washington, USA), Xiaozhong Xu (Tencent, China), and Shan Liu (Tencent, China)

MuGE: Multiple Granularity Edge Detection 25952
Caixia Zhou (Beijing Jiaotong University, China), Yaping Huang (Beijing Jiaotong University, China), Mengyang Pu (North China Electric Power University, China), Qingji Guan (Beijing Jiaotong University, China), Ruoxi Deng (Beijing Jiaotong University, China), and Haibin Ling (Stony Brook University, USA)

KVQ: Kwai Video Quality Assessment for Short-form Videos 25963
Yiting Lu (University of Science and Technology of China), Xin Li (University of Science and Technology of China), Yajing Pei (University of Science and Technology of China), Kun Yuan (Kuaishou Technology), Qizhi Xie (Tsinghua University), Yunpeng Qu (Tsinghua University), Ming Sun (Kuaishou Technology), Chao Zhou (Kuaishou Technology), and Zhibo Chen (University of Science and Technology of China)

Transfer CLIP for Generalizable Image Denoising 25974
Jun Cheng (Huazhong University of Science and Technology, China), Dong Liang (Huazhong University of Science and Technology, China), and Shan Tan (Huazhong University of Science and Technology, China)

Improved Implicit Neural Representation with Fourier Reparameterized Training 25985
Kexuan Shi (University of Electronic Science and Technology of China, China), Xingyu Zhou (University of Electronic Science and Technology of China, China), and Shuhang Gu (University of Electronic Science and Technology of China, China)

Deep Video Inverse Tone Mapping Based on Temporal Clues 25995
Yuyao Ye (Peking University, China), Ning Zhang (Baidu Netdisk), Yang Zhao (Hefei University of Technology, China), Hongbin Cao (Peking University, China; Bytedance Inc.), and Ronggang Wang (Peking University, China)

Boosting Flow-based Generative Super-Resolution Models via Learned Prior 26005
Li-Yuan Tsao (National Tsing Hua University), Yi-Chen Lo (MediaTek Inc.), Chia-Che Chang (MediaTek Inc.), Hao-Wei Chen (National Tsing Hua University), Roy Tseng (MediaTek Inc.), Chien Feng (National Tsing Hua University), and Chun-Yi Lee (National Tsing Hua University)

Look-Up Table Compression for Efficient Image Restoration 26016
Yinglong Li (University of Science and Technology of China, China), Jiacheng Li (University of Science and Technology of China, China), and Zhiwei Xiong (University of Science and Technology of China, China)

Latent Modulated Function for Computational Optimal Continuous Image Representation 26026
Zongyao He (Sun Yat-sen University, China) and Zhi Jin (Sun Yat-sen University, China)

Task-Aware Encoder Control for Deep Video Compression 26036
Xingtong Ge (Beijing Institure of Technology, China), Jixiang Luo (SenseTime Research, China), Xinjie Zhang (The Hong Kong University of Science and Technology, China), Tongda Xu (Institute for AI Industry Research (AIR), Tsinghua University, China), Guo Lu (Shanghai Jiaotong University, China), Dailan He (The Chinese University of Hong Kong, China), Jing Geng (Beijing Institure of Technology, China), Yan Wang (Institute for AI Industry Research (AIR), Tsinghua University, China), Jun Zhang (The Hong Kong University of Science and Technology, China), and Hongwei Qin (SenseTime Research, China)

A Dynamic Kernel Prior Model for Unsupervised Blind Image Super-Resolution 26046
Zhixiong Yang (National University of Defense Technology, China), Jingyuan Xia (National University of Defense Technology, China), Shengxi Li (Beihang University, China), Xinghua Huang (National University of Defense Technology, China), Shuanghui Zhang (National University of Defense Technology, China), Zhen Liu (National University of Defense Technology, China), Yaowen Fu (National University of Defense Technology, China), and Yongxiang Liu (National University of Defense Technology, China)

Zero-Reference Low-Light Enhancement via Physical Quadruple Priors 26057
Wenjing Wang (Peking University), Huan Yang (01.AI), Jianlong Fu (Microsoft Research Asia), and Jiaying Liu (Peking University)

ParamISP: Learned Forward and Inverse ISPs using Camera Parameters 26067
Woohyeok Kim (POSTECH, South Korea), Geonu Kim (POSTECH, South Korea), Junyong Lee (Samsung AI Center Toronto, Canada), Seungyong Lee (POSTECH, South Korea), Seung-Hwan Baek (POSTECH, South Korea), and Sunghyun Cho (POSTECH, South Korea)

FSC: Few-point Shape Completion 26077
Xianzu Wu (Jianghan University, China), Xianfeng Wu (Jianghan University, China), Tianyu Luan (University at Buffalo, USA), Yajing Bai (Jianghan University, China), Zhongyuan Lai (Jianghan University, China), and Junsong Yuan (University at Buffalo, USA)

Generative Latent Coding for Ultra-Low Bitrate Image Compression 26088
Zhaoyang Jia (University of Science and Technology of China), Jiahao Li (Microsoft Research Asia), Bin Li (Microsoft Research Asia), Houqiang Li (University of Science and Technology of China), and Yan Lu (Microsoft Research Asia)

Neural Video Compression with Feature Modulation 26099
Jiahao Li (Microsoft Research Asia), Bin Li (Microsoft Research Asia), and Yan Lu (Microsoft Research Asia)

Driving-Video Dehazing with Non-Aligned Regularization for Safety Assistance 26109
Junkai Fan (Nanjing University of Science and Technology, China), Jiangwei Weng (Nanjing University of Science and Technology, China), Kun Wang (Nanjing University of Science and Technology, China), Yijun Yang (The Hong Kong University of Science and Technology (Guangzhou), China), Jianjun Qian (Nanjing University of Science and Technology, China), Jun Li (Nanjing University of Science and Technology, China), and Jian Yang (Nanjing University of Science and Technology, China)

CFAT: Unleashing Triangular Windows for Image Super-resolution 26120
Abhisek Ray (Senior Research Scholar, Indian Institute of Technology Patna), Gaurav Kumar (MTech, Indian Institute of Technology Patna), and Maheshkumar H. Kolekar (Associate Professor, Indian Institute of Technology Patna)

Zero-Shot Structure-Preserving Diffusion Model for High Dynamic Range Tone Mapping 26130
Ruoxi Zhu (Fudan University, China), Shusong Xu (Alibaba Group, China; Hupan Lab, China), Peiye Liu (Alibaba Group, China; Hupan Lab, China), Sicheng Li (Alibaba Group, China; Hupan Lab, China), Yanheng Lu (Alibaba Group, China; Hupan Lab, China), Dimin Niu (Alibaba Group, China; Hupan Lab, China), Zihao Liu (Alibaba Group, China; Hupan Lab, China), Zihao Meng (Alibaba Group, China; Hupan Lab, China), Zhiyong Li (Alibaba Group, China; Hupan Lab, China), Xinhua Chen (Fudan University, China), and Yibo Fan (Fudan University, China)

Calibrating Multi-modal Representations: A Pursuit of Group Robustness without Annotations. 26140
Chenyu You (Yale University), Yifei Min (Yale University), Weicheng Dai (Yale University), Jasjeet S. Sekhon (Yale University), Lawrence Staib (Yale University), and James S. Duncan (Yale University)

Learn from View Correlation: An Anchor Enhancement Strategy for Multi-view Clustering 26151
Suyuan Liu (National University of Defense Technology, China), Ke Liang (National University of Defense Technology, China), Zhibin Dong (National University of Defense Technology, China), Siwei Wang (Intelligent Game and Decision Lab, China), Xihong Yang (National University of Defense Technology, China), Sihang Zhou (National University of Defense Technology, China), En Zhu (National University of Defense Technology, China), and Xinwang Liu (National University of Defense Technology, China)

Circuit Design and Efficient Simulation of Quantum Inner Product and Empirical Studies of Its Effect on Near-Term Hybrid Quantum-Classic Machine Learning 26162
Hao Xiong (Shanghai Jiao Tong University), Yehui Tang (Shanghai Jiao Tong University), Xinyu Ye (Shanghai Jiao Tong University), and Junchi Yan (Shanghai Jiao Tong University)

Discriminability-Driven Channel Selection for Out-of-Distribution Detection 26171
Yue Yuan (Shandong University, China), Rundong He (Shandong University, China), Yicong Dong (Shandong University, China), Zhongyi Han (King Abdullah University of Science and Technology, Saudi Arabia), and Yilong Yin (Shandong University, China)

Efficient Hyperparameter Optimization with Adaptive Fidelity Identification 26181
Jiantong Jiang (The University of Western Australia, Australia), Zeyi Wen (Hong Kong University of Science and Technology (Guangzhou), China), Atif Mansoor (The University of Western Australia, Australia), and Ajmal Mian (The University of Western Australia, Australia)

Probabilistic Sampling of Balanced K-Means using Adiabatic Quantum Computing 26191
Jan-Nico Zaech (ETH Zurich, Switzerland, INSAIT, Bulgaria), Martin Danelljan (ETH Zurich, Switzerland), Tolga Birdal (Imperial College London, United Kingdom), and Luc Van Gool (ETH Zurich, Switzerland, INSAIT, Bulgaria)

Online Task-Free Continual Generative and Discriminative Learning via Dynamic Cluster Memory 26202
Fei Ye (University of York, UK, Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi) and Adrian G. Bors (University of York, UK, Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi)

Spanning Training Progress: Temporal Dual-Depth Scoring (TDDS) for Enhanced Dataset Pruning 26213
*Xin Zhang (XiDian University, China), Jiawei Du (Agency for Science, Technology and Research (A*STAR), Singapore), Yunsong Li (XiDian University, China), Weiying Xie (XiDian University, China), and Joey Tianyi Zhou (Agency for Science, Technology and Research (A*STAR), Singapore)*

An Aggregation-Free Federated Learning for Tackling Data Heterogeneity 26223
*Yuan Wang (Institute of High Performance Computing, A*STAR, Singapore), Huazhu Fu (Institute of High Performance Computing, A*STAR, Singapore), Renuga Kanagavelu (Institute of High Performance Computing, A*STAR, Singapore), Qingsong Wei (Institute of High Performance Computing, A*STAR, Singapore), Yong Liu (Institute of High Performance Computing, A*STAR, Singapore), and Rick Siow Mong Goh (Institute of High Performance Computing, A*STAR, Singapore)*

POCE: Primal Policy Optimization with Conservative Estimation for Multi-constraint Offline Reinforcement Learning 26233
Jiayi Guan (Tongji University), Li Shen (JD Explore Academy), Ao Zhou (Tongji University), Lusong Li (JD Explore Academy), Han Hu (Beijing Institute of Technology), Xiaodong He (JD Explore Academy), Guang Chen (Tongji University), and Changjun Jiang (Tongji University)

SVDinsTN: A Tensor Network Paradigm for Efficient Structure Search from Regularized Modeling Perspective 26244
Yu-Bang Zheng (Southwest Jiaotong University, China), Xi-Le Zhao (University of Electronic Science and Technology of China, China), Junhua Zeng (Guangdong University of Technology, China), Chao Li (RIKEN, Japan), Qibin Zhao (RIKEN, Japan), Heng-Chao Li (Southwest Jiaotong University, China), and Ting-Zhu Huang (University of Electronic Science and Technology of China, China)

Fine-Grained Bipartite Concept Factorization for Clustering 26254
Chong Peng (Qingdao University, China), Pengfei Zhang (Qingdao University, China), Yongyong Chen (Harbin Institute of Technology (Shenzhen), China), Zhao Kang (University of Electronic Science and Technology of China, China), Chenglizhao Chen (China University of Petroleum (East China)), and Qiang Cheng (University of Kentucky, USA)

Embodied Multi-Modal Agent trained by an LLM from a Parallel TextWorld 26265
Yijun Yang (Southern University of Science and Technology, China), Tianyi Zhou (University of Maryland, College Park, USA), Kanxue Li (Yunnan University, China), Dapeng Tao (Yunnan University, China), Lusong Li (JD Explore Academy), Li Shen (JD Explore Academy), Xiaodong He (JD Explore Academy), Jing Jiang (University of Technology Sydney), and Yuhui Shi (Southern University of Science and Technology, China)

The Mirrored Influence Hypothesis: Efficient Data Influence Estimation by Harnessing Forward Passes 26276
Myeongseob Ko (Virginia Tech), Feiyang Kang (Virginia Tech), Weiyan Shi (Columbia University), Ming Jin (Virginia Tech), Zhou Yu (Columbia University), and Ruoxi Jia (Virginia Tech)

Improved Baselines with Visual Instruction Tuning 26286
Haotian Liu (University of Wisconsin-Madison, USA), Chunyuan Li (Microsoft Research, USA), Yuheng Li (University of Wisconsin-Madison, USA), and Yong Jae Lee (University of Wisconsin-Madison, USA)

Linguistic-Aware Patch Slimming Framework for Fine-grained Cross-Modal Alignment 26297
Zheren Fu (University of Science and Technology of China), Lei Zhang (University of Science and Technology of China), Hou Xia (University of Science and Technology of China), and Zhendong Mao (University of Science and Technology of China)

FlowVQTalker: High-Quality Emotional Talking Face Generation through Normalizing Flow and Quantization 26307
Shuai Tan (Shanghai Jiao Tong University, China), Bin Ji (Shanghai Jiao Tong University, China), and Ye Pan (Shanghai Jiao Tong University, China)

Audio-Visual Segmentation via Unlabeled Frame Exploitation 26318
Jinxiang Liu (Shanghai Jiao Tong University, China), Yikun Liu (Shanghai Jiao Tong University, China), Fei Zhang (Shanghai Jiao Tong University, China), Chen Ju (Shanghai Jiao Tong University, China), Ya Zhang (Shanghai Jiao Tong University, China), and Yanfeng Wang (Shanghai Jiao Tong University, China)

Binding Touch to Everything: Learning Unified Multimodal Tactile Representations 26330
Fengyu Yang (Yale University, USA), Chao Feng (University of Michigan, USA), Ziyang Chen (University of Michigan, USA), Hyoungseob Park (Yale University, USA), Daniel Wang (Yale University, USA), Yiming Dou (University of Michigan, USA), Ziyao Zeng (Yale University, USA), Xien Chen (Yale University, USA), Rit Gangopadhyay (Yale University, USA), Andrew Owens (University of Michigan, USA), and Alex Wong (Yale University, USA)

MoDE: CLIP Data Experts via Clustering 26344
Jiawei Ma (Columbia University), Po-Yao Huang (Meta), Saining Xie (New York University), Shang-Wen Li (Meta), Luke Zettlemoyer (University of Washington), Shih-Fu Chang (Columbia University), Wen-Tau Yih (Meta), and Hu Xu (Meta)

X-MIC: Cross-Modal Instance Conditioning for Egocentric Action Generalization 26354
Anna Kukleva (Meta Reality Lab; Max Planck Institute for Informatics; Saarland Informatics Campus), Fadime Sener (Meta Reality Labs), Edoardo Remelli (Meta Reality Labs), Bugra Tekin (Meta Reality Labs), Eric Sauser (Meta Reality Labs), Bernt Schiele (Max Planck Institute for Informatics, Saarland Informatics Campus), and Shugao Ma (Meta Reality Labs)

PixelLM: Pixel Reasoning with Large Multimodal Model 26364
Zhongwei Ren (Beijing Jiaotong University, China), Zhicheng Huang (University of Science and Technology Beijing, China), Yunchao Wei (Beijing Jiaotong University, China), Yao Zhao (Beijing Jiaotong University, China), Dongmei Fu (University of Science and Technology Beijing, China), Jiashi Feng (ByteDance, Singapore), and Xiaojie Jin (ByteDance, USA)

Probing Synergistic High-Order Interaction in Infrared and Visible Image Fusion 26374
Naishan Zheng (University of Science and Technology of China, China), Man Zhou (University of Science and Technology of China, China), Jie Huang (University of Science and Technology of China, China), Junming Hou (Southeast University, China), Haoying Li (Zhejiang University, China), Yuan Xu (Nanyang Technology University, Singapore), and Feng Zhao (University of Science and Technology of China, China)

The Audio-Visual Conversational Graph: From an Egocentric-Exocentric Perspective 26386
Wenqi Jia (Georgia Institute of Technology), Miao Liu (GenAI, Meta), Hao Jiang (Meta Reality Labs), Ishwarya Ananthabhotla (Meta Reality Labs), James M. Rehg (University of Illinois Urbana-Champaign), Vamsi Krishna Ithapu (Meta Reality Labs), and Ruohan Gao (Meta Reality Labs)

MultiPLY: A Multisensory Object-Centric Embodied Large Language Model in 3D World 26396
Yining Hong (UCLA; MIT-IBM Watson AI Lab), Zishuo Zheng (UMass Amherst), Peihao Chen (SCUT), Yian Wang (UMass Amherst), Junyan Li (UMass Amherst), and Chuang Gan (UMass Amherst; MIT-IBM Watson AI Lab)

GPT4Point: A Unified Framework for Point-Language Understanding and Generation 26407
Zhangyang Qi (The University of Hong Kong), Ye Fang (Fudan University), Zeyi Sun (Shanghai Jiao Tong University), Xiaoyang Wu (The University of Hong Kong), Tong Wu (The Chinese University of Hong Kong), Jiaqi Wang (Shanghai Artificial Intelligence Lab), Dahua Lin (The Chinese University of Hong Kong), and Hengshuang Zhao (The University of Hong Kong)

LL3DA: Visual Interactive Instruction Tuning for Omni-3D Understanding, Reasoning, and Planning 26418
*Sijin Chen (Fudan University), Xin Chen (Tencent PCG), Chi Zhang (Tencent PCG), Mingsheng Li (Fudan University), Gang Yu (Tencent PCG), Hao Fei (National University of Singapore), Hongyuan Zhu (Institute for Infocomm Research (I2R) & Centre for Frontier AI Research (CFAR), A*STAR, Singapore), Jiayuan Fan (Fudan University), and Tao Chen (Fudan University)*

Unified-IO 2: Scaling Autoregressive Multimodal Models with Vision, Language, Audio, and Action 26429
Jiasen Lu (Allen Institute for AI), Christopher Clark (Allen Institute for AI), Sangho Lee (Allen Institute for AI), Zichen Zhang (Allen Institute for AI), Savya Khosla (University of Illinois Urbana-Champaign), Ryan Marten (University of Illinois Urbana-Champaign), Derek Hoiem (University of Illinois Urbana-Champaign), and Aniruddha Kembhavi (Allen Institute for AI)

SHAP-EDITOR: Instruction-Guided Latent 3D Editing in Seconds 26446
Minghao Chen (University of Oxford), Junyu Xie (University of Oxford), Iro Laina (University of Oxford), and Andrea Vedaldi (University of Oxford)

Learning to Visually Localize Sound Sources from Mixtures without Prior Source Knowledge ... 26457
Dongjin Kim (Kyung Hee University), Sung Jin Um (Kyung Hee University), Sangmin Lee (University of Illinois Urbana-Champaign), and Jung Uk Kim (Kyung Hee University)

Bring Event into RGB and LiDAR: Hierarchical Visual-Motion Fusion for Scene Flow 26467
Hanyu Zhou (Huazhong University of Science and Technology), Yi Chang (Huazhong University of Science and Technology), and Zhiwei Shi (Huazhong University of Science and Technology)

Dispel Darkness for Better Fusion: A Controllable Visual Enhancer based on Cross-modal Conditional Adversarial Learning 26477
Hao Zhang (Wuhan University), Linfeng Tang (Wuhan University), Xinyu Xiang (Wuhan University), Xuhui Zuo (Wuhan University), and Jiayi Ma (Wuhan University)

Unraveling Instance Associations: A Closer Look for Audio-Visual Segmentation 26487
Yuanhong Chen (University of Adelaide), Yuyuan Liu (University of Adelaide), Hu Wang (University of Adelaide), Fengbei Liu (University of Adelaide), Chong Wang (University of Adelaide), Helen Frazer (St Vincent's Hospital Melbourne), and Gustavo Carneiro (University of Surrey)

DMR: Decomposed Multi-Modality Representations for Frames and Events Fusion in Visual Reinforcement Learning 26498
Haoran Xu (Sun Yat-sen University, China; Peng Cheng Laboratory, China), Peixi Peng (Peking University, China; Peng Cheng Laboratory, China), Guang Tan (Sun Yat-sen University, China), Yuan Li (Academy of Military Sciences, China), Xinhai Xu (Academy of Military Sciences, China), and Yonghong Tian (Peking University, China; Peng Cheng Laboratory, China)

Text-Guided Variational Image Generation for Industrial Anomaly Detection and Segmentation. 26509
Mingyu Lee (Chung-Ang University, Korea) and Jongwon Choi (Chung-Ang University, Korea)

Tactile-Augmented Radiance Fields 26519
Yiming Dou (University of Michigan), Fengyu Yang (Yale University), Yi Liu (University of Michigan), Antonio Loquercio (University of California, Berkeley), and Andrew Owens (University of Michigan)

LION: Empowering Multimodal Large Language Model with Dual-Level Visual Knowledge 26530
Gongwei Chen (Harbin Institute of Technology, Shenzhen, China), Leyang Shen (Harbin Institute of Technology, Shenzhen, China), Rui Shao (Harbin Institute of Technology, Shenzhen, China), Xiang Deng (Harbin Institute of Technology, Shenzhen, China), and Liqiang Nie (Harbin Institute of Technology, Shenzhen, China)

SDSTrack: Self-Distillation Symmetric Adapter Learning for Multi-Modal Visual Object Tracking 26541
Xiaojun Hou (Zhejiang University), Jiazheng Xing (Zhejiang University), Yijie Qian (Zhejiang University), Yaowei Guo (Zhejiang University), Shuo Xin (Zhejiang University), Junhao Chen (Zhejiang University), Kai Tang (Zhejiang University), Mengmeng Wang (Zhejiang University), Zhengkai Jiang (Youtu Lab, Tencent), Liang Liu (Huzhou Institute, Zhejiang University), and Yong Liu (Zhejiang University)

Exploring the Transferability of Visual Prompting for Multimodal Large Language Models 26552
Yichi Zhang (Tsinghua University, China), Yinpeng Dong (Tsinghua University, China), Siyuan Zhang (Tsinghua University, China), Tianzan Min (Tsinghua University, China), Hang Su (Tsinghua University, China), and Jun Zhu (Tsinghua University, China)

Mask Grounding for Referring Image Segmentation 26563
Yong Xien Chng (Tsinghua University, China), Henry Zheng (Tsinghua University, China), Yizeng Han (Tsinghua University, China), Xuchong Qiu (Bosch Corporate Research, Germany), and Gao Huang (Tsinghua University, China)

OneLLM: One Framework to Align All Modalities with Language 26574
Jiaming Han (The Chinese University of Hong Kong, China), Kaixiong Gong (The Chinese University of Hong Kong, China), Yiyuan Zhang (The Chinese University of Hong Kong, China), Jiaqi Wang (Shanghai Artificial Intelligence Laboratory, China), Kaipeng Zhang (Shanghai Artificial Intelligence Laboratory, China), Dahua Lin (The Chinese University of Hong Kong, China; Shanghai Artificial Intelligence Laboratory, China), Yu Qiao (Shanghai Artificial Intelligence Laboratory, China), Peng Gao (Shanghai Artificial Intelligence Laboratory, China), and Xiangyu Yue (The Chinese University of Hong Kong, China)

EmoVIT: Revolutionizing Emotion Insights with Visual Instruction Tuning 26586
Hongxia Xie (Jilin University), Chu-Jun Peng (National Taiwan University), Yu-Wen Tseng (National Taiwan University), Hung-Jen Chen (National Yang Ming Chiao Tung University), Chan-Feng Hsu (National Yang Ming Chiao Tung University), Hong-Han Shuai (National Yang Ming Chiao Tung University), and Wen-Huang Cheng (National Taiwan University)

ModaVerse: Efficiently Transforming Modalities with LLMs 26596
Xinyu Wang (The University of Adelaide), Bohan Zhuang (Monash University), and Qi Wu (The University of Adelaide)

PromptKD: Unsupervised Prompt Distillation for Vision-Language Models 26607
Zheng Li (Nankai University, China), Xiang Li (Nankai University, China), Xinyi Fu (Ant Group, China), Xin Zhang (Nankai University, China), Weiqiang Wang (Ant Group, China), Shuo Chen (RIKEN, Japan), and Jian Yang (Nankai University, China)

Dynamic Prompt Optimizing for Text-to-Image Generation 26617
Wenyi Mo (Renmin University of China, China), Tianyu Zhang (Du Xiaoman Technology, China), Yalong Bai (Du Xiaoman Technology, China), Bing Su (Renmin University of China, China), Ji-Rong Wen (Renmin University of China, China), and Qing Yang (Du Xiaoman Technology, China)

Domain Prompt Learning with Quaternion Networks 26627
Qinglong Cao (Shanghai Jiao Tong University), Zhengqin Xu (Shanghai Jiao Tong University), Yuntian Chen (Eastern Institute of Technology, Ningbo), Chao Ma (Shanghai Jiao Tong University), and Xiaokang Yang (Shanghai Jiao Tong University)

ViT-Lens: Towards Omni-modal Representations 26637
Weixian Lei (National University of Singapore, Singapore), Yixiao Ge (Tencent, China), Kun Yi (Tencent, China), Jianfeng Zhang (National University of Singapore, Singapore), Difei Gao (National University of Singapore, Singapore), Dylan Sun (Tencent, China), Yuying Ge (Tencent, China), Ying Shan (Tencent, China), and Mike Zheng Shou (National University of Singapore, Singapore)

Rotated Multi-Scale Interaction Network for Referring Remote Sensing Image Segmentation 26648
Sihan Liu (Xiamen University), Yiwei Ma (Xiamen University), Xiaoqing Zhang (Xiamen University), Haowei Wang (Xiamen University), Jiayi Ji (Xiamen University), Xiaoshuai Sun (Xiamen University), and Rongrong Ji (Xiamen University)

Cyclic Learning for Binaural Audio Generation and Localization 26659
Zhaojian Li (Northwestern Polytechnical University, China), Bin Zhao (Northwestern Polytechnical University, China), and Yuan Yuan (Northwestern Polytechnical University, China)

Learning to Rematch Mismatched Pairs for Robust Cross-Modal Retrieval 26669
Haochen Han (Xi'an Jiaotong University, China), Qinghua Zheng (Xi'an Jiaotong University, China), Guang Dai (SGIT AI Lab, China), Minnan Luo (Xi'an Jiaotong University, China), and Jingdong Wang (Baidu Inc, China)

VILA: On Pre-training for Visual Language Models 26679
Ji Lin (MIT), Hongxu Yin (NVIDIA), Wei Ping (NVIDIA), Pavlo Molchanov (NVIDIA), Mohammad Shoeybi (NVIDIA), and Song Han (MIT)

A Picture is Worth More Than 77 Text Tokens: Evaluating CLIP-Style Models on Dense Captions 26690
Jack Urbanek (FAIR, Meta), Florian Bordes (FAIR, Meta; Mila; Universite de Montreal), Pietro Astolfi (FAIR, Meta), Mary Williamson (FAIR, Meta), Vasu Sharma (FAIR, Meta), and Adriana Romero-Soriano (FAIR, Meta: Universite de Montreal; McGill University; Canada CIFAR AI)

How to Configure Good In-Context Sequence for Visual Question Answering 26700
Li Li (Key Laboratory of New Generation Artificial Intelligence Technology and Its Interdisciplinary Applications (Southeast University), Ministry of Education, China), Jiawei Peng (Key Laboratory of New Generation Artificial Intelligence Technology and Its Interdisciplinary Applications (Southeast University), Ministry of Education, China), Huiyi Chen (Key Laboratory of New Generation Artificial Intelligence Technology and Its Interdisciplinary Applications (Southeast University), Ministry of Education, China), Chongyang Gao (Northwestern University), and Xu Yang (Key Laboratory of New Generation Artificial Intelligence Technology and Its Interdisciplinary Applications (Southeast University), Ministry of Education, China)

CrossMAE: Cross-Modality Masked Autoencoders for Region-Aware Audio-Visual Pre-Training 26711
Yuxin Guo (Institute of Automation, Chinese Academy of Sciences), Siyang Sun (Alibaba Group), Shuailei Ma (Northeastern University), Kecheng Zheng (Ant Group), Xiaoyi Bao (Institute of Automation, Chinese Academy of Sciences), Shijie Ma (Institute of Automation, Chinese Academy of Sciences), Wei Zou (Institute of Automation, Chinese Academy of Sciences), and Yun Zheng (Alibaba Group)

Modality-Collaborative Test-Time Adaptation for Action Recognition 26722
Baochen Xiong (Institute of Automation, Chinese Academy of Sciences; Peng Cheng Laboratory), Xiaoshan Yang (Institute of Automation, Chinese Academy of Sciences; Peng Cheng Laboratory), Yaguang Song (Peng Cheng Laboratory), Yaowei Wang (Peng Cheng Laboratory), and Changsheng Xu (Institute of Automation, Chinese Academy of Sciences; Peng Cheng Laboratory)

T-VSL: Text-Guided Visual Sound Source Localization in Mixtures 26732
Tanvir Mahmud (The University of Texas at Austin), Yapeng Tian (The University of Texas at Dallas), and Diana Marculescu (The University of Texas at Austin)

UniBind: LLM-Augmented Unified and Balanced Representation Space to Bind Them All 26742
Yuanhuiyi Lyu (The Hong Kong University of Science and Technology (Guangzhou)), Xu Zheng (The Hong Kong University of Science and Technology (Guangzhou)), Jiazhou Zhou (The Hong Kong University of Science and Technology (Guangzhou)), and Lin Wang (The Hong Kong University of Science and Technology (Guangzhou))

Monkey: Image Resolution and Text Label Are Important Things for Large Multi-modal Models 26753
Zhang Li (Huazhong University of Science and Technology, China), Biao Yang (Huazhong University of Science and Technology, China), Qiang Liu (Kingsoft Office, China), Zhiyin Ma (Huazhong University of Science and Technology, China), Shuo Zhang (Huazhong University of Science and Technology, China), Jingxu Yang (Kingsoft Office, China), Yabo Sun (Kingsoft Office, China), Yuliang Liu (Huazhong University of Science and Technology, China), and Xiang Bai (Huazhong University of Science and Technology, China)

Rethinking Multi-view Representation Learning via Distilled Disentangling 26764
Guanzhou Ke (Beijing Jiaotong University), Bo Wang (Institute of Automation, Chinese Academy of Sciences), Xiaoli Wang (Nanjing University of Science and Technology), and Shengfeng He (Singapore Management University)

Causal Mode Multiplexer: A Novel Framework for Unbiased Multispectral Pedestrian Detection..... 26774
Taeheon Kim (KAIST, South Korea), Sebin Shin (KAIST, South Korea), Youngjoon Yu (KAIST, South Korea), Hak Gu Kim (Chung-Ang University, South Korea), and Yong Man Ro (KAIST, South Korea)

Image-Text Co-Decomposition for Text-Supervised Semantic Segmentation 26784
Ji-Jia Wu (National Taiwan University, Taiwan), Andy Chia-Hao Chang (National Yang Ming Chiao Tung University, Taiwan), Chieh-Yu Chuang (National Yang Ming Chiao Tung University, Taiwan), Chun-Pei Chen (National Yang Ming Chiao Tung University, Taiwan), Yu-Lun Liu (National Yang Ming Chiao Tung University, Taiwan), Min-Hung Chen (NVIDIA, Taiwan), Hou-Ning Hu (Mediatek.inc, Taiwan), Yung-Yu Chuang (National Taiwan University, Taiwan), and Yen-Yu Lin (National Yang Ming Chiao Tung University, Taiwan)

Mirasol3B: A Multimodal Autoregressive Model for Time-Aligned and Contextual Modalities ... 26794
AJ Piergiovanni (Google Deepmind), Isaac Noble (Google Research), Dahun Kim (Google Deepmind), Michael S. Ryoo (Google Deepmind), Victor Gomes (Google Research), and Anelia Angelova (Google Deepmind)

Efficient Vision-Language Pre-training by Cluster Masking 26805
Zihao Wei (University of Michigan), Zixuan Pan (University of Michigan), and Andrew Owens (University of Michigan)

MeLFusion: Synthesizing Music from Image and Language Cues using Diffusion Models 26816
Sanjoy Chowdhury (University of Maryland College Park, USA), Sayan Nag (University of Toronto, Canada), K J Joseph (Adobe Research, India), Balaji Vasan Srinivasan (Adobe Research, India), and Dinesh Manocha (University of Maryland College Park, USA)

Weakly Misalignment-free Adaptive Feature Alignment for UAVs-based Multimodal Object Detection 26826
Chen Chen (National University of Defense Technology, China), Jiahao Qi (National University of Defense Technology, China), Xingyue Liu (National University of Defense Technology, China), Kangcheng Bin (National University of Defense Technology, China), Ruigang Fu (National University of Defense Technology, China), Xikun Hu (National University of Defense Technology, China), and Ping Zhong (National University of Defense Technology, China)

DiVAS: Video and Audio Synchronization with Dynamic Frame Rates 26836
Clara Fernandez-Labrador (Disney Research|Studios), Mertcan Akçay (ETH Zurich), Eitan Abecassis (Disney Entertainment and ESPN Technology), Joan Massich (Disney Research|Studios), and Christopher Schroers (Disney Research|Studios)

Querying as Prompt: Parameter-Efficient Learning for Multimodal Language Model 26845
Tian Liang (Zhejiang University, China), Jing Huang (Zhejiang University, China), Ming Kong (Zhejiang University, China), Luyuan Chen (Beijing Information Science and Technology University, China), and Qiang Zhu (Zhejiang University, China)

SonicVisionLM: Playing Sound with Vision Language Models 26856
Zhifeng Xie (Shanghai University, China), Shengye Yu (Shanghai University, China), Qile He (Shanghai University, China), and Mengtian Li (Shanghai University, China)

Embracing Unimodal Aleatoric Uncertainty for Robust Multimodal Fusion 26866
Zixian Gao (University of Electronic Science and Technology of China), Xun Jiang (University of Electronic Science and Technology of China), Xing Xu (University of Electronic Science and Technology of China; Tongji University), Fumin Shen (University of Electronic Science and Technology of China), Yujie Li (Kyushu Institute of Technology), and Heng Tao Shen (University of Electronic Science and Technology of China; Tongji University)

C3Net: Compound Conditioned ControlNet for Multimodal Content Generation 26876
Juntao Zhang (The Hong Kong University of Science and Technology, China), Yuehuai Liu (The Hong Kong University of Science and Technology, China), Yu-Wing Tai (Dartmouth College, USA), and Chi-Keung Tang (The Hong Kong University of Science and Technology, China)

Composed Video Retrieval via Enriched Context and Discriminative Embeddings 26886
Omkar Thawakar (MBZUAI), Muzammal Naseer (MBZUAI), Rao Muhammad Anwer (MBZUAI/AALTO), Salman Khan (MBZUAI), Michael Felsberg (Linköping University), Mubarak Shah (University of Central Florida), and Fahad Shahbaz Khan (MBZUAI)

Looking Similar, Sounding Different: Leveraging Counterfactual Cross-Modal Pairs for Audiovisual Representation Learning 26897
Nikhil Singh (Massachusetts Institute of Technology, USA), Chih-Wei Wu (Netflix, USA), Iroro Orife (Netflix, USA), and Mahdi Kalayeh (Netflix, USA)

Anchor-based Robust Finetuning of Vision-Language Models 26909
Jinwei Han (Wuhan University), Zhiwen Lin (YouTu Lab, Tencent), Zhongyisun Sun (YouTu Lab, Tencent), Yingguo Gao (YouTu Lab, Tencent), Ke Yan (YouTu Lab, Tencent), Shouhong Ding (YouTu Lab, Tencent), Yuan Gao (Wuhan University), and Gui-Song Xia (Wuhan University)

Event-based Visible and Infrared Fusion via Multi-task Collaboration 26919
Mengyue Geng (Peking University, China), Lin Zhu (Beijing Institute of Technology, China), Lizhi Wang (Beijing Institute of Technology, China), Wei Zhang (Peng Cheng Laboratory, China), Ruiqin Xiong (Peking University, China), and Yonghong Tian (Peking University, China)

Prompt Learning via Meta-Regularization 26930
Jinyoung Park (Korea University, South Korea), Juyeon Ko (Korea University, South Korea), and Hyunwoo J. Kim (Korea University, South Korea)

Knowledge-Enhanced Dual-stream Zero-shot Composed Image Retrieval 26941
Yucheng Suo (ReLER, CCAI, Zhejiang University, China), Fan Ma (ReLER, CCAI, Zhejiang University, China), Linchao Zhu (ReLER, CCAI, Zhejiang University, China), and Yi Yang (ReLER, CCAI, Zhejiang University, China)

Contextual Augmented Global Contrast for Multimodal Intent Recognition 26953
Kaili Sun (Wuhan University, China), Zhiwen Xie (Central China Normal University, China), Mang Ye (Wuhan University, China), and Huyin Zhang (Wuhan University, China)

MRFS: Mutually Reinforcing Image Fusion and Segmentation 26964
Hao Zhang (Wuhan University), Xuhui Zuo (Wuhan University), Jie Jiang (Tencent), Chunchao Guo (Tencent), and Jiayi Ma (Wuhan University)

POPDG: Popular 3D Dance Generation with PopDanceSet 26974
Zhenye Luo (Beijing Normal University, China), Min Ren (Beijing Normal University, China), Xuecai Hu (Beijing Normal University, China), Yongzhen Huang (Beijing Normal University, China), and Li Yao (Beijing Normal University, China)

How to Make Cross Encoder a Good Teacher for Efficient Image-Text Retrieval? 26984
Yuxin Chen (Institute of Automation, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China; Tencent PCG, China), Zongyang Ma (Institute of Automation, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China; Tencent PCG, China), Ziqi Zhang (Institute of Automation, Chinese Academy of Sciences, China), Zhongang Qi (Tencent PCG, China), Chunfeng Yuan (Institute of Automation, Chinese Academy of Sciences, China), Bing Li (Institute of Automation, Chinese Academy of Sciences, China), Junfu Pu (Tencent PCG, China), Ying Shan (Tencent PCG, China), Xiaojuan Qi (The University of Hong Kong, China), and Weiming Hu (Institute of Automation, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China; ShanghaiTech University, China)

Active Prompt Learning in Vision Language Models 26994
Jihwan Bang (KAIST), Sumyeong Ahn (Michigan State University), and Jae-Gil Lee (KAIST)

Descriptor and Word Soups: Overcoming the Parameter Efficiency Accuracy Tradeoff for Out-of-Distribution Few-shot Learning 27005
Christopher Liao (Boston University), Theodoros Tsiligkaridis (MIT Lincoln Laboratory), and Brian Kulis (Boston University)

Text-IF: Leveraging Semantic Text Guidance for Degradation-Aware and Interactive Image Fusion 27016
Xunpeng Yi (Wuhan University), Han Xu (Wuhan University), Hao Zhang (Wuhan University), Linfeng Tang (Wuhan University), and Jiayi Ma (Wuhan University)

Hallucination Augmented Contrastive Learning for Multimodal Large Language Model 27026
Chaoya Jiang (Peking University), Haiyang Xu (Alibaba Group), Mengfan Dong (Peking University), Jiaxing Chen (Peking University), Wei Ye (Peking University), Ming Yan (Alibaba Group), Qinghao Ye (Alibaba Group), Ji Zhang (Alibaba Group), Fei Huang (Alibaba Group), and Shikun Zhang (Peking University)

Beyond Text: Frozen Large Language Models in Visual Signal Comprehension 27037
Lei Zhu (Peking University, China), Fangyun Wei (Microsoft Research Asia, China), and Yanye Lu (Peking University, China)

Learning Spatial Features from Audio-Visual Correspondence in Egocentric Videos 27048
Sagnik Majumder (University of Texas at Austin; Meta), Ziad Al-Halah (University of Utah), and Kristen Grauman (University of Texas at Austin; Meta)

ES³: Evolving Self-Supervised Learning of Robust Audio-Visual Speech Representations 27059
Yuanhang Zhang (Institute of Computing Technology, Chinese Academy of Sciences), Shuang Yang (Institute of Computing Technology, Chinese Academy of Sciences), Shiguang Shan (Institute of Computing Technology, Chinese Academy of Sciences), and Xilin Chen (Institute of Computing Technology, Chinese Academy of Sciences)

PortraitBooth: A Versatile Portrait Model for Fast Identity-preserved Personalization 27070
Xu Peng (Xiamen University, China), Junwei Zhu (Tencent, China), Boyuan Jiang (Tencent, China), Ying Tai (Nanjing University, China), Donghao Luo (Tencent, China), Jiangning Zhang (Tencent, China), Wei Lin (Xiamen University, China), Taisong Jin (Xiamen University, China), Chengjie Wang (Tencent, China), and Rongrong Ji (Xiamen University, China)

ULIP-2: Towards Scalable Multimodal Pre-training for 3D Understanding 27081
Le Xue (Salesforce AI Research, USA), Ning Yu (Salesforce AI Research, USA), Shu Zhang (Salesforce AI Research, USA), Artemis Panagopoulou (Salesforce AI Research, USA; University of Pennsylvania, USA), Junnan Li (Salesforce AI Research, Singapore), Roberto Martín-Martín (University of Texas at Austin, USA), Jiajun Wu (Stanford University, USA), Caiming Xiong (Salesforce AI Research, USA), Ran Xu (Salesforce AI Research, USA), Juan Carlos Niebles (Salesforce AI Research, USA; Stanford University, USA), and Silvio Savarese (Salesforce AI Research, USA; Stanford University, USA)

AVFF: Audio-Visual Feature Fusion for Video Deepfake Detection 27092
Trevine Oorloff (University of Maryland - College Park), Surya Koppisetti (Reality Defender Inc.), Nicolò Bonettini (Reality Defender Inc.), Divyaraj Solanki (Reality Defender Inc.), Ben Colman (Reality Defender Inc.), Yaser Yacoob (University of Maryland - College Park), Ali Shahriyari (Reality Defender Inc.), and Gaurav Bharaj (Reality Defender Inc.)

Language-aware Visual Semantic Distillation for Video Question Answering 27103
Bo Zou (Tsinghua University, China), Chao Yang (Shanghai AI Laboratory, China), Yu Qiao (Shanghai AI Laboratory, China), Chengbin Quan (Tsinghua University, China), and Youjian Zhao (Tsinghua University, China)

PerceptionGPT: Effectively Fusing Visual Perception into LLM 27114
Renjie Pi (Hong Kong University of Science and Technology), Lewei Yao (Hong Kong University of Science and Techology), Jiahui Gao (the University of Hong Kong), Jipeng Zhang (Hong Kong University of Science and Technology), and Tong Zhang (Hong Kong University of Science and Technology)

Cooperation Does Matter: Exploring Multi-Order Bilateral Relations for Audio-Visual Segmentation 27124
Qi Yang (University of Chinese Academy of Sciences, China), Xing Nie (Institute of Automation Chinese Academy of Sciences, China), Tong Li (Meituan, China), Pengfei Gao (Meituan, China), Ying Guo (Meituan, China), Cheng Zhen (Meituan, China), Pengfei Yan (Meituan, China), and Shiming Xiang (Institute of Automation Chinese Academy of Sciences, China)

MV-Adapter: Multimodal Video Transfer Learning for Video Text Retrieval 27134
Xiaojie Jin (Bytedance Inc., USA), Bowen Zhang (Tsinghua University, China), Weibo Gong (Bytedance Inc., China), Kai Xu (Bytedance Inc., China), Xueqing Deng (Bytedance Inc., USA), Peng Wang (Bytedance Inc., USA), Zhao Zhang (Hefei University of Technology, China), Xiaohui Shen (Bytedance Inc., USA), and Jiashi Feng (Bytedance Inc., Singapore)

Draw Step by Step: Reconstructing CAD Construction Sequences from Point Clouds via Multimodal Diffusion. 27144
Weijian Ma (Fudan University, China), Shuaiqi Chen (Fudan University, China), Yunzhong Lou (Fudan University, China), Xueyang Li (Fudan University, China), and Xiangdong Zhou (Fudan University, China)

AV-RIR: Audio-Visual Room Impulse Response Estimation 27154
Anton Ratnarajah (University of Maryland, College Park), Sreyan Ghosh (University of Maryland, College Park), Sonal Kumar (University of Maryland, College Park), Purva Chiniya (University of Maryland, College Park), and Dinesh Manocha (University of Maryland, College Park)

Link-Context Learning for Multimodal LLMs 27166
Yan Tai (Ningbo Institute of Digital Twin, Eastern Institute of Technology, Ningbo, China), Weichen Fan (HyperGAI), Zhao Zhang (Sensetime), and Ziwei Liu (S-Lab, Nanyang Technological University)

Unveiling the Power of Audio-Visual Early Fusion Transformers with Dense Interactions through Masked Modeling 27176
Shentong Mo (Carnegie Mellon University) and Pedro Morgado (University of Wisconsin-Madison)

Noisy-Correspondence Learning for Text-to-Image Person Re-identification 27187
*Yang Qin (Sichuan University, China), Yingke Chen (Northumbria University, United Kingdom), Dezhong Peng (Sichuan University, China), Xi Peng (Sichuan University, China), Joey Tianyi Zhou (A*STAR, Singapore), and Peng Hu (Sichuan University, China)*

Mind Artist: Creating Artistic Snapshots with Human Thought 27197
Jiaxuan Chen (Zhejiang University, China), Yu Qi (Zhejiang University, China), Yueming Wang (Zhejiang University, China), and Gang Pan (Zhejiang University, China)

VTQA: Visual Text Question Answering via Entity Alignment and Cross-Media Reasoning 27208
Kang Chen (Harbin Institute of Technology, China) and Xiangqian Wu (Harbin Institute of Technology, China)

THRONE: An Object-based Hallucination Benchmark for the Free-form Generations of Large Vision-Language Models 27218
Prannay Kaul (VGG, University of Oxf), Zhizhong Li (AWS AI Labs), Hao Yang (AWS AI Labs), Yonatan Dukler (AWS AI Labs), Ashwin Swaminathan (AWS AI Labs), C. J. Taylor (AWS AI Labs), and Stefano Soatto (AWS AI Labs)

Data-Efficient Multimodal Fusion on a Single GPU 27229
Noël Vouitsis (Layer 6 AI), Zhaoyan Liu (Layer 6 AI), Satya Krishna Gorti (Layer 6 AI), Valentin Villecroze (Layer 6 AI), Jesse C. Cresswell (Layer 6 AI), Guangwei Yu (Layer 6 AI), Gabriel Loaiza-Ganem (Layer 6 AI), and Maksims Volkovs (Layer 6 AI)

SoundingActions: Learning How Actions Sound from Narrated Egocentric Videos 27242
Changan Chen (University of Texas at Austin, USA), Kumar Ashutosh (University of Texas at Austin, USA), Rohit Girdhar (Meta AI, USA), David Harwath (University of Texas at Austin, USA), and Kristen Grauman (University of Texas at Austin, USA)

Accept the Modality Gap: An Exploration in the Hyperbolic Space 27253
Sameera Ramasinghe (Amazon), Violetta Shevchenko (Amazon), Gil Avraham (Amazon), and Ajanthan Thalaiyasingam (Amazon)

DiffSal: Joint Audio and Video Learning for Diffusion Saliency Prediction 27263
Junwen Xiong (Northwestern Polytechnical University, China), Peng Zhang (Northwestern Polytechnical University, China), Tao You (Northwestern Polytechnical University, China), Chuanyue Li (Northwestern Polytechnical University, China), Wei Huang (Nanchang University, China), and Yufei Zha (Northwestern Polytechnical University, China)

DiPrompT: Disentangled Prompt Tuning for Multiple Latent Domain Generalization in Federated Learning 27274
Sikai Bai (The Hong Kong University of Science and Technology, Hong Kong SAR), Jie Zhang (The Hong Kong Polytechnic University, Hong Kong SAR), Song Guo (The Hong Kong University of Science and Technology, Hong Kong SAR), Shuaicheng Li (Sensetime Research, China), Jingcai Guo (The Hong Kong Polytechnic University, Hong Kong SAR), Jun Hou (Sensetime Research, China), Tao Han (The Hong Kong University of Science and Technology, Hong Kong SAR), and Xiaocheng Lu (The Hong Kong University of Science and Technology, Hong Kong SAR)

Probabilistic Speech-Driven 3D Facial Motion Synthesis: New Benchmarks, Methods, and Applications 27284
Karren D. Yang (Apple), Anurag Ranjan (Apple), Jen-Hao Rick Chang (Apple), Raviteja Vemulapalli (Apple), and Oncel Tuzel (Apple)

DIEM: Decomposition-Integration Enhancing Multimodal Insights 27294
Xinyi Jiang (Zhejiang University, China), Guoming Wang (Zhejiang University, China), Junhao Guo (Zhejiang University), Juncheng Li (Zhejiang University), Wenqiao Zhang (Zhejiang University), Rongxing Lu (University of New Brunswick), and Siliang Tang (Zhejiang University)

MAFA: Managing False Negatives for Vision-Language Pre-training 27304
Jaeseok Byun (Seoul National University, the Republic of Korea), Dohoon Kim (Seoul National University, the Republic of Korea), and Taesup Moon (Seoul National University, the Republic of Korea)

AV2AV: Direct Audio-Visual Speech to Audio-Visual Speech Translation with Unified Audio-Visual Speech Representation 27315
Jeongsoo Choi (KAIST, South Korea), Se Jin Park (KAIST, South Korea), Minsu Kim (KAIST, South Korea), and Yong Man Ro (KAIST, South Korea)

Enhancing Multimodal Cooperation via Sample-level Modality Valuation 27328
Yake Wei (Renmin University of China, China), Ruoxuan Feng (Renmin University of China, China), Zihe Wang (Renmin University of China, China), and Di Hu (Renmin University of China, China)

Diff-BGM: A Diffusion Model for Video Background Music Generation 27338
Sizhe Li (Peking University, China), Yiming Qin (Peking University, China), Minghang Zheng (Peking University, China), Xin Jin (Beijing Electronic Science and Technology Institute), and Yang Liu (Peking University, China)

SaCo Loss: Sample-wise Affinity Consistency for Vision-Language Pre-training 27348
Sitong Wu (The Chinese University of Hong Kong (CUHK)), Haoru Tan (The University of Hong Kong (HKU)), Zhuotao Tian (SmartMore), Yukang Chen (The Chinese University of Hong Kong (CUHK)), Xiaojuan Qi (The University of Hong Kong (HKU)), and Jiaya Jia (The Chinese University of Hong Kong (CUHK), SmartMore)

MoPE-CLIP: Structured Pruning for Efficient Vision-Language Models with Module-wise Pruning Error Metric 27360
Haokun Lin (University of Chinese Academy of Sciences (UCAS), CRIPAC & MAIS, Institute of Automation, Chinese Academy of Sciences, City University of Hong Kong), Haoli Bai (Huawei Noah's Ark Lab), Zhili Liu (The Hong Kong University of Science and Technology, Huawei Noah's Ark Lab), Lu Hou (Huawei Noah's Ark Lab), Muyi Sun (CRIPAC & MAIS, Institute of Automation, Chinese Academy of Sciences), Linqi Song (City University of Hong Kong, City University of Hong Kong Shenzhen Research Institute), Ying Wei (Nanyang Technological University), and Zhenan Sun (CRIPAC & MAIS, Institute of Automation, Chinese Academy of Sciences)

Mitigating Noisy Correspondence by Geometrical Structure Consistency Learning 27371
Zihua Zhao (Shanghai Jiao Tong University, China), Mengxi Chen (Shanghai Jiao Tong University, China), Tianjie Dai (Shanghai Jiao Tong University, China), Jiangchao Yao (Shanghai Jiao Tong University, China), Bo Han (Hong Kong Baptist University), Ya Zhang (Shanghai Jiao Tong University, China), and Yanfeng Wang (Shanghai Jiao Tong University, China)

DetCLIPv3: Towards Versatile Generative Open-vocabulary Object Detection 27381
Lewei Yao (Hong Kong University of Science and Technology, Hong Kong), Renjie Pi (Hong Kong University of Science and Technology, Hong Kong), Jianhua Han (Huawei Noah's Ark Lab, China), Xiaodan Liang (Shenzhen Campus of Sun Yat-Sen University), Hang Xu (Huawei Noah's Ark Lab, China), Wei Zhang (Huawei Noah's Ark Lab, China), Zhenguo Li (Huawei Noah's Ark Lab, China), and Dan Xu (Hong Kong University of Science and Technology, Hong Kong)

Leveraging Cross-Modal Neighbor Representation for Improved CLIP Classification 27392
Chao Yi (Nanjing University, China), Lu Ren (Nanjing University, China), De-Chuan Zhan (Nanjing University, China), and Han-Jia Ye (Nanjing University, China)

OmniVec2 - A Novel Transformer based Network for Large Scale Multimodal and Multitask Learning 27402
Siddharth Srivastava (Typeface) and Gaurav Sharma (Typeface)

CoDi-2: In-Context, Interleaved, and Interactive Any-to-Any Generation 27415
Zineng Tang (UC Berkeley), Ziyi Yang (Microsoft Azure AI), Mahmoud Khademi (Microsoft Azure AI), Yang Liu (Microsoft Azure AI), Chenguang Zhu (Zoom), and Mohit Bansal (UNC Chapel Hill)

Differentiable Information Bottleneck for Deterministic Multi-view Clustering 27425
Xiaoqiang Yan (Zhengzhou University, China), Zhixiang Jin (Zhengzhou University, China), Fengshou Han (Zhengzhou University, China), and Yangdong Ye (Zhengzhou University, China)

A Study of Dropout-Induced Modality Bias on Robustness to Missing Video Frames for Audio-Visual Speech Recognition 27435
Yusheng Dai (University of Science and Techlogy of China), Hang Chen (University of Science and Technology of China), Jun Du (University of Science and Technology of China), Ruoyu Wang (University of Science and Technology of China), Shihao Chen (University of Science and Technology of China), Haotian Wang (University of Science and Technology of China), and Chin-Hui Lee (Georgia Institute of Technology)

Multimodal Representation Learning by Alternating Unimodal Adaptation 27446
Xiaohui Zhang (UNC-Chapel Hill), Jaehong Yoon (UNC-Chapel Hill), Mohit Bansal (UNC-Chapel Hill), and Huaxiu Yao (UNC-Chapel Hill)

View-Category Interactive Sharing Transformer for Incomplete Multi-View Multi-Label Learning 27457
Shilong Ou (Beijing University of Posts and Telecommunications, China), Zhe Xue (Beijing University of Posts and Telecommunications, China), Yawen Li (Beijing University of Posts and Telecommunications, China), Meiyu Liang (Beijing University of Posts and Telecommunications, China), Yuanqiang Cai (Beijing University of Posts and Telecommunications, China), and Junjiang Wu (Beijing University of Posts and Telecommunications, China)

Scalable 3D Registration via Truncated Entry-wise Absolute Residuals 27467
Tianyu Huang (The Chinese University of Hong Kong), Liangzu Peng (University of Pennsylvania), Rene Vidal (University of Pennsylvania), and Yun-Hui Liu (The Chinese University of Hong Kong)

Partial-to-Partial Shape Matching with Geometric Consistency 27478
Viktoria Ehm (Technical University of Munich, Germany), Maolin Gao (Technical University of Munich, Germany), Paul Roetzer (University of Bonn, Germany), Marvin Eisenberger (Technical University of Munich, Germany), Daniel Cremers (Technical University of Munich, Germany), and Florian Bernard (University of Bonn, Germany)

Towards Robust Learning to Optimize with Theoretical Guarantees 27488
Qingyu Song (The Chinese University of Hong Kong, China), Wei Lin (The Chinese University of Hong Kong, China), Juncheng Wang (Hong Kong Baptist University, China), and Hong Xu (The Chinese University of Hong Kong, China)

From Variance to Veracity: Unbundling and Mitigating Gradient Variance in Differentiable Bundle Adjustment Layers 27497
Swaminathan Gurumurthy (Carnegie Mellon University), Karnik Ram (TU Munich), Bingqing Chen (Bosch Center for Artificial Intelligence), Zachary Manchester (Carnegie Mellon University), and Zico Kolter (Carnegie Mellon University, Bosch Center for Artificial Intelligence)

DIMAT: Decentralized Iterative Merging-And-Training for Deep Learning Models 27507
Nastaran Saadati (Iowa State University, USA), Minh Pham (New York University, USA), Nasla Saleem (Iowa State University, USA), Joshua R. Waite (Iowa State University, USA), Aditya Balu (Iowa State University, USA), Zhanong Jiang (Iowa State University, USA), Chinmay Hegde (New York University, USA), and Soumik Sarkar (Iowa State University, USA)

Ink Dot-Oriented Differentiable Optimization for Neural Image Halftoning 27518
Hao Jiang (Peking University), Bingfeng Zhou (Peking University), and Yadong Mu (Peking University)

Are Conventional SNNs Really Efficient? A Perspective from Network Quantization 27528
Guobin Shen (Institute of Automation, Chinese Academy of Sciences), Dongcheng Zhao (Institute of Automation, Chinese Academy of Sciences), Tenglong Li (Institute of Automation, Chinese Academy of Sciences), Jindong Li (Institute of Automation, Chinese Academy of Sciences), and Yi Zeng (Institute of Automation, Chinese Academy of Sciences)

FedMef: Towards Memory-efficient Federated Dynamic Pruning 27538
Hong Huang (City University of Hong Kong), Weiming Zhuang (Sony AI), Chen Chen (Sony AI), and Lingjuan Lyu (Sony AI)

SD4Match: Learning to Prompt Stable Diffusion Model for Semantic Matching 27548
Xinghui Li (University of Oxford), Jingyi Lu (University of Hong Kong), Kai Han (University of Hong Kong), and Victor Adrian Prisacariu (University of Oxford)

Purified and Unified Steganographic Network 27559
Guobiao Li (Fudan University, China), Sheng Li (Fudan University, China), Zicong Luo (Fudan University, China), Zhenxing Qian (Fudan University, China), and Xinpeng Zhang (Fudan University, China)

Learned Lossless Image Compression based on Bit Plane Slicing 27569
Zhe Zhang (Wuhan University, China), Huairui Wang (Wuhan University, China), Zhenzhong Chen (Wuhan University, China), and Shan Liu (Tencent Media Lab, USA)

Towards Calibrated Multi-label Deep Neural Networks 27579
Jiacheng Cheng (University of California San Diego, USA) and Nuno Vasconcelos (University of California San Diego, USA)

Improving Generalization via Meta-Learning on Hard Samples 27590
Nishant Jain (Google Research India), Arun S. Suggala (Google Research India), and Pradeep Shenoy (Google Research India)

Learning with Structural Labels for Learning with Noisy Labels 27600
Noo-ri Kim (Sungkyunkwan University), Jin-Seop Lee (Sungkyunkwan University), and Jee-Hyong Lee (Sungkyunkwan University)

DiffuseMix: Label-Preserving Data Augmentation with Diffusion Models 27611
Khawar Islam (FloppyDisk.AI), Muhammad Zaigham Zaheer (Mohamed bin Zayed University of Artificial Intelligence), Arif Mahmood (Information Technology University, Punjab), and Karthik Nandakumar (Mohamed bin Zayed University of Artificial Intelligence)

Improving Out-of-Distribution Generalization in Graphs via Hierarchical Semantic Environments 27621
Yinhua Piao (Seoul National University), Sangseon Lee (Seoul National University), Yijingxiu Lu (Seoul National University), and Sun Kim (Seoul National University, AIGENDRUG Co., Ltd.)

Patch2Self2: Self-supervised Denoising on Coresets via Matrix Sketching 27631
Shreyas Fadnavis (Johnson and Johnson), Agniva Chowdhury (Oak Ridge National Laboratory), Joshua Batson (Anthropic), Petros Drineas (Purdue University), and Eleftherios Garyfallidis (Indiana University Bloomington)

G-FARS: Gradient-Field-based Auto-Regressive Sampling for 3D Part Grouping 27642
Junfeng Cheng (Imperial College London) and Tania Stathaki (Imperial College London)

Decompose-and-Compose: A Compositional Approach to Mitigating Spurious Correlation 27652
Fahimeh Hosseini Noohdani (Sharif University of Technology, Iran), Parsa Hosseini (Sharif University of Technology, Iran), Aryan Yazdan Parast (Sharif University of Technology, Iran), Hamidreza Yaghoubi Araghi (Sharif University of Technology, Iran), and Mahdieh Soleymani Baghshah (Sharif University of Technology, Iran)

SkySense: A Multi-Modal Remote Sensing Foundation Model Towards Universal Interpretation for Earth Observation Imagery 27662
Xin Guo (Ant Group, China), Jiangwei Lao (Ant Group, China), Bo Dang (Wuhan University, China), Yingying Zhang (Ant Group, China), Lei Yu (Ant Group, China), Lixiang Ru (Ant Group, China), Liheng Zhong (Ant Group, China), Ziyuan Huang (Ant Group, China), Kang Wu (Wuhan University, China), Dingxiang Hu (Ant Group, China), Huimei He (Ant Group, China), Jian Wang (Ant Group, China), Jingdong Chen (Ant Group, China), Ming Yang (Ant Group, China), Yongjun Zhang (Wuhan University, China), and Yansheng Li (Wuhan University, China)

Building Bridges across Spatial and Temporal Resolutions: Reference-Based Super-Resolution via Change Priors and Conditional Diffusion Model 27674
Runmin Dong (Tsinghua University, China), Shuai Yuan (The University of Hong Kong, China), Bin Luo (Tsinghua University, China), Mengxuan Chen (Tsinghua University, China), Jinxiao Zhang (Tsinghua University, China), Lixian Zhang (National Supercomputing Center in Shenzhen, China), Weijia Li (Sun Yat-Sen University, China), Juepeng Zheng (Sun Yat-Sen University, China), and Haohuan Fu (Tsinghua University, China)

SatSynth: Augmenting Image-Mask Pairs through Diffusion Models for Aerial Semantic Segmentation 27685
Aysim Toker (Technical University of Munich), Marvin Eisenberger (Technical University of Munich), Daniel Cremers (Technical University of Munich), and Laura Leal-Taixé (NVIDIA)

S2MAE: A Spatial-Spectral Pretraining Foundation Model for Spectral Remote Sensing Data 27696
Xuyang Li (Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), Danfeng Hong (Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), and Jocelyn Chanussot (Univ. Grenoble Alpes, France)

Poly Kernel Inception Network for Remote Sensing Detection 27706
Xinhao Cai (Nanjing University of Science and Technology), Qiuxia Lai (Communication University of China), Yuwei Wang (Nanjing University of Science and Technology), Wenguan Wang (Zhejiang University), Zeren Sun (Nanjing University of Science and Technology), and Yazhou Yao (Nanjing University of Science and Technology)

Learning without Exact Guidance: Updating Large-scale High-resolution Land Cover Maps from Low-resolution Historical Labels 27717
Zhuohong Li (Wuhan University, China), Wei He (Wuhan University, China), Jiepan Li (Wuhan University, China), Fangxiao Lu (Wuhan University, China), and Hongyan Zhang (Wuhan University, China)

3D Building Reconstruction from Monocular Remote Sensing Images with Multi-level Supervisions 27728
Weijia Li (Sun Yat-sen University), Haote Yang (Shanghai AI Laboratory), Zhenghao Hu (Sun Yat-sen University), Juepeng Zheng (Sun Yat-sen University), Gui-Song Xia (Wuhan University), and Conghui He (Shanghai AI Laboratory)

Content-Adaptive Non-Local Convolution for Remote Sensing Pansharpening 27738
Yule Duan (University of Electronic Science and Technology of China, China), Xiao Wu (University of Electronic Science and Technology of China, China), Haoyu Deng (University of Electronic Science and Technology of China, China), and Liang-Jian Deng (University of Electronic Science and Technology of China, China)

SG-BEV: Satellite-Guided BEV Fusion for Cross-View Semantic Segmentation 27748
Junyan Ye (Sun Yat-Sen University, China), Qiyan Luo (Sun Yat-Sen University, China), Jinhua Yu (Sun Yat-Sen University, China), Huaping Zhong (SenseTime Research, China), Zhimeng Zheng (Zhejiang University, China), Conghui He (Shanghai AI Laboratory, China), and Weijia Li (Sun Yat-Sen University, China)

DiffCast: A Unified Framework via Residual Diffusion for Precipitation Nowcasting 27758
Demin Yu (Harbin Institute of Technology Shenzhen, China), Xutao Li (Harbin Institute of Technology Shenzhen, China), Yunming Ye (Harbin Institute of Technology Shenzhen, China), Baoquan Zhang (Harbin Institute of Technology Shenzhen, China), Chuyao Luo (Harbin Institute of Technology Shenzhen, China), Kuai Dai (Harbin Institute of Technology Shenzhen, China), Rui Wang (Shenzhen Meteorological Bureau), and Xunlai Chen (Shenzhen Meteorological Bureau)

MoCha-Stereo: Motif Channel Attention Network for Stereo Matching 27768
Ziyang Chen (Guizhou University), Wei Long (Guizhou University), He Yao (Guizhou University), Yongjun Zhang (Guizhou University), Bingshu Wang (Northwest Polytechnical University), Yongbin Qin (Guizhou University), and Jia Wu (Guizhou University)

PBWR: Parametric-Building-Wireframe Reconstruction from Aerial LiDAR Point Clouds 27778
Shangfeng Huang (University of Calgary, Canada), Ruisheng Wang (University of Calgary, Canada), Bo Guo (Guangdong University of Technology, China), and Hongxin Yang (University of Calgary, Canada)

Multi-modal Learning for Geospatial Vegetation Forecasting 27788
Vitus Benson (Max-Planck-Institute for Biogeochemistry, Germany), Claire Robin (Max-Planck-Institute for Biogeochemistry, Germany), Christian Requena-Mesa (Max-Planck-Institute for Biogeochemistry, Germany), Lazaro Alonso (Max-Planck-Institute for Biogeochemistry), Nuno Carvalhais (Max-Planck-Institute for Biogeochemistry, Germany), José Cortés (Max-Planck-Institute for Biogeochemistry, Germany), Zhihan Gao (Hong Kong University of Science and Technology, Hong Kong), Nora Linscheid (Max-Planck-Institute for Biogeochemistry, Germany), Mélanie Weynants (Max-Planck-Institute for Biogeochemistry, Germany), and Markus Reichstein (Max-Planck-Institute for Biogeochemistry)

Relational Matching for Weakly Semi-Supervised Oriented Object Detection 27800
Wenhao Wu (City University of Hong Kong, China), Hau-San Wong (City University of Hong Kong, China), Si Wu (South China University of Technology, China), and Tianyou Zhang (South China University of Technology, China)

Rethinking Transformers Pre-training for Multi-Spectral Satellite Imagery 27811
Mubashir Noman (Mohamed bin Zayed University of Artificial Intelligence, United Arab Emirates), Muzammal Naseer (Mohamed bin Zayed University of Artificial Intelligence, United Arab Emirates), Hisham Cholakkal (Mohamed bin Zayed University of Artificial Intelligence, United Arab Emirates), Rao Muhammad Anwer (Mohamed bin Zayed University of Artificial Intelligence, United Arab Emirates), Salman Khan (Mohamed bin Zayed University of Artificial Intelligence, United Arab Emirates), and Fahad Shahbaz Khan (Mohamed bin Zayed University of Artificial Intelligence, United Arab Emirates)

Unmixing Diffusion for Self-Supervised Hyperspectral Image Denoising 27820
Haijin Zeng (IMEC-UGent, Belgium), Jiezhang Cao (ETH Zurich, Switzerland), Kai Zhang (Nanjing University, China), Yongyong Chen (Harbin Institute of Technology, China), Hiep Luong (IMEC-UGent, Belgium), and Wilfried Philips (IMEC-UGent, Belgium)

GeoChat: Grounded Large Vision-Language Model for Remote Sensing 27831
Kartik Kuckreja (MBZUAI), Muhammad Sohail Danish (MBZUAI), Muzammal Naseer (MBZUAI), Abhijit Das (BITS-Pilani, Hyderabad), Salman Khan (MBZUAI), and Fahad Shahbaz Khan (MBZUAI)

Parameter Efficient Self-Supervised Geospatial Domain Adaptation 27841
Linus Scheibenreif (University of St.Gallen, Switzerland), Michael Mommert (Stuttgart University of Applied Sciences, Germany), and Damian Borth (University of St.Gallen, Switzerland)

Bridging Remote Sensors with Multisensor Geospatial Foundation Models 27852
Boran Han (Amazon Web Services), Shuai Zhang (Amazon Web Services), Xingjian Shi (Boson AI), and Markus Reichstein (Amazon Web Services\; Max Planck Institute for Biogeochemistry)

CLIP-Driven Open-Vocabulary 3D Scene Graph Generation via Cross-Modality Contrastive Learning 27863
Lianggangxu Chen (East China Normal University, China), Xuejiao Wang (East China Normal University, China), Jiale Lu (East China Normal University, China), Shaohui Lin (East China Normal University, China), Changbo Wang (East China Normal University, China), and Gaoqi He (East China Normal University, China)

Learnable Earth Parser: Discovering 3D Prototypes in Aerial Scans 27874
Romain Loiseau (LIGM, Ecole des Ponts, Univ Gustave Eiffel, CNRS, France and Univ Gustave Eiffel, IGN, ENSG, LASTIG, France), Elliot Vincent (LIGM, Ecole des Ponts, Univ Gustave Eiffel, CNRS, France and INRIA, Paris, France), Mathieu Aubry (LIGM, Ecole des Ponts, Univ Gustave Eiffel, CNRS, France), and Loic Landrieu (LIGM, Ecole des Ponts, Univ Gustave Eiffel, CNRS, France and Univ Gustave Eiffel, IGN, ENSG, LASTIG, France)

Semantics, Distortion, and Style Matter: Towards Source-free UDA for Panoramic Segmentation 27885
Xu Zheng (The Hong Kong University of Science and Technology (Guangzhou), China), Pengyuan Zhou (Aarhus University, Denmark), Athanasios V. Vasilakos (University of Agder, Norway), and Lin Wang (The Hong Kong University of Science and Technology (Guangzhou), China)

Geometrically-driven Aggregation for Zero-shot 3D Point Cloud Understanding 27896
Guofeng Mei (Fondazione Bruno Kessler), Luigi Riz (Fondazione Bruno Kessler), Yiming Wang (Fondazione Bruno Kessler), and Fabio Poiesi (Fondazione Bruno Kessler)

SAM-6D: Segment Anything Model Meets Zero-Shot 6D Object Pose Estimation 27906
Jiehong Lin (DexForce Co. Ltd., Shenzhen, China), Lihua Liu (South China University of Technology, China), Dekun Lu (South China University of Technology, China), and Kui Jia (The Chinese University of Hong Kong, Shenzhen, China)

Construct to Associate: Cooperative Context Learning for Domain Adaptive Point Cloud Segmentation 27917
Guangrui Li (University of Technology Sydney)

Multi-Task Dense Prediction via Mixture of Low-Rank Experts 27927
Yuqi Yang (Nankai University, China), Peng-Tao Jiang (vivo Mobile Communication Co., Ltd, China), Qibin Hou (Nankai University, China), Hao Zhang (vivo Mobile Communication Co., Ltd, China), Jinwei Chen (vivo Mobile Communication Co., Ltd, China), and Bo Li (vivo Mobile Communication Co., Ltd, China)

OED: Towards One-stage End-to-End Dynamic Scene Graph Generation 27938
Guan Wang (Peking University, China), Zhimin Li (Tencent Inc, China), Qingchao Chen (Peking University, China), and Yang Liu (Peking University, China)

OMG-Seg: Is One Model Good Enough For All Segmentation? 27948
Xiangtai Li (Nanyang Technological University), Haobo Yuan (Nanyang Technological University), Wei Li (Nanyang Technological University), Henghui Ding (Nanyang Technological University), Size Wu (Nanyang Technological University), Wenwei Zhang (Nanyang Technological University), Yining Li (Shanghai Artificial Intelligence Laboratory), Kai Chen (Shanghai Artificial Intelligence Laboratory), and Chen Change Loy (Nanyang Technological University)

DiffusionMTL: Learning Multi-Task Denoising Diffusion Model from Partially Annotated Data . 27960
Hanrong Ye (HKUST) and Dan Xu (HKUST)

Bilateral Adaptation for Human-Object Interaction Detection with Occlusion-Robustness 27970
Guangzhi Wang (National University of Singapore), Yangyang Guo (National University of Singapore), Ziwei Xu (National University of Singapore), and Mohan Kankanhalli (National University of Singapore)

CurveCloudNet: Processing Point Clouds with 1D Structure 27981
Colton Stearns (Stanford University), Alex Fu (Stanford University), Jiateng Liu (University of Illinois Urbana-Champaign), Jeong Joon Park (University of Michigan), Davis Rempe (NVIDIA), Despoina Paschalidou (Stanford University), and Leonidas J. Guibas (Stanford University)

VCoder: Versatile Vision Encoders for Multimodal Large Language Models 27992
Jitesh Jain (Georgia Tech), Jianwei Yang (Microsoft Research), and Humphrey Shi (Georgia Tech\; Picsart AI Research)

Amodal Ground Truth and Completion in the Wild 28003
Guanqi Zhan (University of Oxford, UK), Chuanxia Zheng (University of Oxford, UK), Weidi Xie (Shanghai Jiao Tong University, China), and Andrew Zisserman (University of Oxford, UK)

Living Scenes: Multi-object Relocalization and Reconstruction in Changing 3D Environments 28014
Liyuan Zhu (Stanford University), Shengyu Huang (ETH Zurich), Konrad Schindler (ETH Zurich), and Iro Armeni (Stanford University)

Single Domain Generalization for Crowd Counting 28025
Zhuoxuan Peng (The Hong Kong University of Science and Technology, Hong Kong) and S.-H. Gary Chan (The Hong Kong University of Science and Technology, Hong Kong)

LTA-PCS: Learnable Task-Agnostic Point Cloud Sampling 28035
Jiaheng Liu (Beihang University), Jianhao Li (Beihang University), Kaisiyuan Wang (Baidu Inc.), Hongcheng Guo (Beihang University), Jian Yang (Alibaba DAMO Academy), Junran Peng (Chinese Academy of Sciences), Ke Xu (Beihang University), Xianglong Liu (Beihang University), and Jinyang Guo (Beihang University)

Prompt3D: Random Prompt Assisted Weakly-Supervised 3D Object Detection 28046
Xiaohong Zhang (Nanjing University), Huisheng Ye (Nanjing University), Jingwen Li (Nanjing University), Qinyu Tang (Nanjing University), Yuanqi Li (Nanjing University), Yanwen Guo (Nanjing University), and Jie Guo (Nanjing University)

No More Ambiguity in 360° Room Layout via Bi-Layout Estimation 28056
Yu-Ju Tsai (University of California Merced, USA), Jin-Cheng Jhang (National Tsing Hua University, Taiwan), Jingjing Zheng (Amazon, USA), Wei Wang (Amazon, USA), Albert Y. C. Chen (Amazon, USA), Min Sun (Amazon, USA), Cheng-Hao Kuo (Amazon, USA), and Ming-Hsuan Yang (University of California Merced, USA)

Semantic Line Combination Detector 28066
Jinwon Ko (Korea University), Dongkwon Jin (Korea University), and Chang-Su Kim (Korea University)

From Pixels to Graphs: Open-Vocabulary Scene Graph Generation with Vision-Language Models 28076
Rongjie Li (School of Information Science and Technology), Songyang Zhang (Shanghai AI Laboratory), Dahua Lin (Shanghai AI Laboratory), Kai Chen (Shanghai AI Laboratory), and Xuming He (ShanghaiTech University; Shanghai Engineering Research Center of Intelligent Vision and Imaging)

PanoContext-Former: Panoramic Total Scene Understanding with a Transformer 28087
Yuan Dong (Alibaba Group, China), Chuan Fang (Hong Kong University of Science and Technology, Hong Kong), Liefeng Bo (Alibaba Group, China), Zilong Dong (Alibaba Group, China), and Ping Tan (Hong Kong University of Science and Technology, Hong Kong)

DiffAssemble: A Unified Graph-Diffusion Model for 2D and 3D Reassembly 28098
Gianluca Scarpellini (Istituto Italiano di Tecnologia), Stefano Fiorini (Istituto Italiano di Tecnologia), Francesco Giuliari (Istituto Italiano di Tecnologia), Pietro Moreiro (Istituto Italiano di Tecnologia), and Alessio Del Bue (Istituto Italiano di Tecnologia)

ProMotion: Prototypes As Motion Learners 28109
Yawen Lu (Purdue University), Dongfang Liu (Rochester Institute of Technology), Qifan Wang (Meta AI), Cheng Han (Rochester Institute of Technology), Yiming Cui (Tiktok), Zhiwen Cao (Purdue University), Xueling Zhang (Rochester Institute of Technology), Yingjie Victor Chen (Purdue University), and Heng Fan (University of North Texas)

HUNTER: Unsupervised Human-centric 3D Detection via Transferring Knowledge from Synthetic Instances to Real Scenes 28120
Yichen Yao (ShanghaiTech University), Zimo Jiang (ShanghaiTech University; University of Chinese Academy of Sciences; Innovation Academy for Microsatellites, Chinese Academy of Sciences), Yujing Sun (The University of Hong Kong), Zhencai Zhu (Innovation Academy for Microsatellites, Chinese Academy of Sciences), Xinge Zhu (The Chinese University of Hong Kong), Runnan Chen (The University of Hong Kong), and Yuexin Ma (ShanghaiTech University)

Rethinking the Up-Sampling Operations in CNN-based Generative Network for Generalizable Deepfake Detection 28130
Chuangchuang Tan (Beijing Jiaotong University), Yao Zhao (Beijing Jiaotong University), Shikui Wei (Beijing Jiaotong University), Guanghua Gu (Yanshan University), Ping Liu (University of Nevada), and Yunchao Wei (Beijing Jiaotong University)

Shadows Don't Lie and Lines Can't Bend! Generative Models don't know Projective Geometry...for now 28140
Ayush Sarkar (University of Illinois Urbana Champaign), Hanlin Mai (University of Illinois Urbana Champaign), Amitabh Mahapatra (University of Illinois Urbana Champaign), Svetlana Lazebnik (University of Illinois Urbana Champaign), D.A. Forsyth (University of Illinois Urbana Champaign), and Anand Bhattad (Toyota Technological Institute at Chicago)

Text Grouping Adapter: Adapting Pre-trained Text Detector for Layout Analysis 28150
Tianci Bi (Xi'an Jiaotong University), Xiaoyi Zhang (Microsoft Research Asia), Zhizheng Zhang (Microsoft Research Asia), Wenxuan Xie (Microsoft Research Asia), Cuiling Lan (Microsoft Research Asia), Yan Lu (Microsoft Research Asia), and Nanning Zheng (Xi'an Jiaotong University)

Groupwise Query Specialization and Quality-Aware Multi-Assignment for Transformer-based Visual Relationship Detection 28160
Jongha Kim (Korea University, South Korea), Jihwan Park (Korea University, South Korea), Jinyoung Park (Korea University, South Korea), Jinyoung Kim (Korea University, South Korea), Sehyung Kim (Korea University, South Korea), and Hyunwoo J. Kim (Korea University, South Korea)

CoralSCOP: Segment any COral Image on this Planet 28170
Ziqiang Zheng (The Hong Kong University of Science and Technology, China), Haixin Liang (The Hong Kong University of Science and Technology, China), Binh-Son Hua (Trinity College Dublin, Ireland), Yue Him Wong (Shenzhen University, China), Put Ang Jr (The Chinese University of Hong Kong, China), Apple Pui Yi Chui (The Chinese University of Hong Kong, China), and Sai-Kit Yeung (The Hong Kong University of Science and Technology, China)

Going Beyond Multi-Task Dense Prediction with Synergy Embedding Models 28181
Huimin Huang (Zhejiang University), Yawen Huang (Jarvis Research Center, Tencent YouTu Lab), Lanfen Lin (Zhejiang University), Ruofeng Tong (Zhejiang University, Zhejiang Lab), Yen-Wei Chen (Ritsumeikan University), Hao Zheng (Jarvis Research Center, Tencent YouTu Lab), Yuexiang Li (Guangxi Medical University), and Yefeng Zheng (Jarvis Research Center, Tencent YouTu Lab)

Disentangled Pre-training for Human-Object Interaction Detection 28191
Zhuolong Li (South China University of Technology), Xingao Li (South China University of Technology), Changxing Ding (South China University of Technology), and Xiangmin Xu (South China University of Technology)

Osprey: Pixel Understanding with Visual Instruction Tuning 28202
Yuqian Yuan (Zhejiang University, China), Wentong Li (Zhejiang University, China), Jian Liu (Ant Group, China), Dongqi Tang (Ant Group, China), Xinjie Luo (Zhejiang University, China), Chi Qin (Microsoft, China), Lei Zhang (The HongKong Polytechnical University, China), and Jianke Zhu (Zhejiang University, China)

Discovering Syntactic Interaction Clues for Human-Object Interaction Detection 28212
Jinguo Luo (Harbin Institute of Technology, Shenzhen, China), Weihong Ren (Harbin Institute of Technology, Shenzhen, China), Weibo Jiang (Harbin Institute of Technology, Shenzhen, China), Xi'ai Chen (Shenyang Institute of Automation, Chinese Academy of Sciences, China), Qiang Wang (Shenyang University, China), Zhi Han (Shenyang Institute of Automation, Chinese Academy of Sciences, China), and Honghai Liu (Harbin Institute of Technology, Shenzhen, China)

Flattening the Parent Bias: Hierarchical Semantic Segmentation in the Poincaré Ball 28223
Simon Weber (Technical University of Munich), Barış Zöngür (Technical University of Munich), Nikita Araslanov (Technical University of Munich), and Daniel Cremers (Technical University of Munich)

HiKER-SGG: Hierarchical Knowledge Enhanced Robust Scene Graph Generation 28233
Ce Zhang (Carnegie Mellon University), Simon Stepputtis (Carnegie Mellon University), Joseph Campbell (Carnegie Mellon University), Katia Sycara (Carnegie Mellon University), and Yaqi Xie (Carnegie Mellon University)

Hierarchical Intra-modal Correlation Learning for Label-free 3D Semantic Segmentation 28244
Xin Kang (University of Science and Technology of China, China), Lei Chu (Microsoft Research Asia, China), Jiahao Li (Microsoft Research Asia, China), Xuejin Chen (University of Science and Technology of China, China), and Yan Lu (Microsoft Research Asia, China)

FreePoint: Unsupervised Point Cloud Instance Segmentation 28254
Zhikai Zhang (Wuhan University, China), Jian Ding (Wuhan University, China; KAUST, Saudi Arabia), Li Jiang (The Chinese University of Hong Kong, Shenzhen, China), Dengxin Dai (Huawei Zurich Research Center), and Guisong Xia (Wuhan University, China)

GoodSAM: Bridging Domain and Capacity Gaps via Segment Anything Model for Distortion-aware Panoramic Semantic Segmentation 28264
Weiming Zhang (HKUST(GZ), China), Yexin Liu (HKUST(GZ), China), Xu Zheng (HKUST(GZ), China), and Lin Wang (HKUST(GZ), China)

MaskClustering: View Consensus based Mask Graph Clustering for Open-Vocabulary 3D Instance Segmentation 28274
Mi Yan (Peking University, China), Jiazhao Zhang (Peking University, China), Yan Zhu (Peking University, China), and He Wang (Peking University, China)

ECoDepth: Effective Conditioning of Diffusion Models for Monocular Depth Estimation 28285
Suraj Patni (Indian Institute of Technology Delhi), Aradhye Agarwal (Indian Institute of Technology Delhi), and Chetan Arora (Indian Institute of Technology Delhi)

Physical Property Understanding from Language-Embedded Feature Fields 28296
Albert J. Zhai (University of Illinois at Urbana-Champaign), Yuan Shen (University of Illinois at Urbana-Champaign), Emily Y. Chen (University of Illinois at Urbana-Champaign), Gloria X. Wang (University of Illinois at Urbana-Champaign), Xinlei Wang (University of Illinois at Urbana-Champaign), Sheng Wang (University of Illinois at Urbana-Champaign), Kaiyu Guan (University of Illinois at Urbana-Champaign), and Shenlong Wang (University of Illinois at Urbana-Champaign)

LLM4SGG: Large Language Models for Weakly Supervised Scene Graph Generation 28306
Kibum Kim (KAIST), Kanghoon Yoon (KAIST), Jaehyeong Jeon (KAIST), Yeonjun In (KAIST), Jinyoung Moon (ETRI), Donghyun Kim (Korea University), and Chanyoung Park (KAIST)

DSGG: Dense Relation Transformer for an End-to-end Scene Graph Generation 28317
Zeeshan Hayder (Data61-CSIRO, Australia) and Xuming He (ShanghaiTech University)

OTE: Exploring Accurate Scene Text Recognition Using One Token 28327
Jianjun Xu (University of Science and Technology of China, China), Yuxin Wang (University of Science and Technology of China, China), Hongtao Xie (University of Science and Technology of China, China), and Yongdong Zhang (University of Science and Technology of China, China)

SemCity: Semantic Scene Generation with Triplane Diffusion 28337
Jumin Lee (KAIST, South Korea), Sebin Lee (KAIST, South Korea), Changho Jo (Neosapience, Inc., South Korea), Woobin Im (KAIST, South Korea), Juhyeong Seon (KAIST, South Korea), and Sung-Eui Yoon (KAIST, South Korea)

Advancing Saliency Ranking with Human Fixations: Dataset, Models and Benchmarks 28348
Bowen Deng (University of Nottingham, UK), Siyang Song (University of Leicester, UK), Andrew P. French (University of Nottingham, UK), Denis Schluppeck (University of Nottingham, UK), and Michael P. Pound (University of Nottingham, UK)

Choose What You Need: Disentangled Representation Learning for Scene Text Recognition, Removal and Editing 28358
Boqiang Zhang (University of Science and Technology of China), Hongtao Xie (University of Science and Technology of China), Zuan Gao (University of Science and Technology of China), and Yuxin Wang (University of Science and Technology of China)

Leveraging Predicate and Triplet Learning for Scene Graph Generation 28369
Jiankai Li (Beihang University, China; Shanghai Artificial Intelligence Laboratory, China), Yunhong Wang (Beihang University, China), Xiefan Guo (Beihang University, China), Ruijie Yang (Beihang University, China), and Weixin Li (Beihang University, China; Shanghai Artificial Intelligence Laboratory, China)

Regressor-Segmenter Mutual Prompt Learning for Crowd Counting 28380
Mingyue Guo (University of Chinese Academy of Sciences, China), Li Yuan (Peking University, China), Zhaoyi Yan (Pengcheng Lab, China), Binghui Chen (Alibaba, China), Yaowei Wang (Pengcheng Lab, China), and Qixiang Ye (University of Chinese Academy of Sciences, China)

Learning from Observer Gaze: Zero-Shot Attention Prediction Oriented by Human-Object Interaction Recognition 28390
Yuchen Zhou (Sun Yat-sen University, China), Linkai Liu (Sun Yat-sen University, China), and Chao Gou (Sun Yat-sen University, China)

SG-PGM: Partial Graph Matching Network with Semantic Geometric Fusion for 3D Scene Graph Alignment and Its Downstream Tasks 28401
Yaxu Xie (German Research Center for Artificial Intelligence), Alain Pagani (German Research Center for Artificial Intelligence), and Didier Stricker (German Research Center for Artificial Intelligence)

Open-Vocabulary Semantic Segmentation with Image Embedding Balancing 28412
Xiangheng Shan (Huazhong University of Science and Technology, China), Dongyue Wu (Huazhong University of Science and Technology, China), Guilin Zhu (Huazhong University of Science and Technology, China), Yuanjie Shao (Huazhong University of Science and Technology, China), Nong Sang (Huazhong University of Science and Technology, China), and Changxin Gao (Huazhong University of Science and Technology, China)

Bridging the Synthetic-to-Authentic Gap: Distortion-Guided Unsupervised Domain Adaptation for Blind Image Quality Assessment 28422
Aobo Li (Xidian University, China), Jinjian Wu (Xidian University, China), Yongxu Liu (Xidian University, China), and Leida Li (Xidian University, China)

Robust Distillation via Untargeted and Targeted Intermediate Adversarial Samples 28432
*Junhao Dong (Nanyang Technological University, Singapore; A*STAR, Singapore), Piotr Koniusz (CSIRO's Data61, Australia; Australian National University, Australia), Junxi Chen (Sun Yat-sen University, China), Z. Jane Wang (University of British Columbia, Canada), and Yew-Soon Ong (Nanyang Technological University, Singapore; A*STAR, Singapore)*

Class Incremental Learning with Multi-Teacher Distillation 28443
Haitao Wen (University of Electronic Science and Technology of China, China), Lili Pan (University of Electronic Science and Technology of China, China), Yu Dai (University of Electronic Science and Technology of China, China), Heqian Qiu (University of Electronic Science and Technology of China, China), Lanxiao Wang (University of Electronic Science and Technology of China, China), Qingbo Wu (University of Electronic Science and Technology of China, China), and Hongliang Li (University of Electronic Science and Technology of China, China)

Large Language Models are Good Prompt Learners for Low-Shot Image Classification 28453
Zhaoheng Zheng (University of Southern California, USA), Jingmin Wei (University of Southern California, USA), Xuefeng Hu (University of Southern California, USA), Haidong Zhu (University of Southern California, USA), and Ram Nevatia (University of Southern California, USA)

Consistent Prompting for Rehearsal-Free Continual Learning 28463
Zhanxin Gao (Sun Yat-sen University, China), Jun Cen (The Hong Kong University of Science and Technology, China), and Xiaobin Chang (Sun Yat-sen University, China)

Tuning Stable Rank Shrinkage: Aiming at the Overlooked Structural Risk in Fine-tuning 28474
Sicong Shen (Beihang University, China), Yang Zhou (Beihang University, China), Bingzheng Wei (Xiaomi Corporation, China), Eric I-Chao Chang (Taiwan Artificial Intelligence Foundation, China), and Yan Xu (Beihang University, China)

Coherent Temporal Synthesis for Incremental Action Segmentation 28485
Guodong Ding (National University of Singapore), Hans Golong (National University of Singapore), and Angela Yao (National University of Singapore)

FCS: Feature Calibration and Separation for Non-Exemplar Class Incremental Learning 28495
Qiwei Li (Peking University, China), Yuxin Peng (Peking University, China), and Jiahuan Zhou (Peking University, China)

DeIL: Direct-and-Inverse CLIP for Open-World Few-Shot Learning 28505
Shuai Shao (Zhejiang Lab, China), Yu Bai (China University of Petroleum (East China), China), Yan Wang (Beihang University, China), Baodi Liu (China University of Petroleum (East China), China), and Yicong Zhou (University of Macau, China)

Understanding and Improving Source-free Domain Adaptation from a Theoretical Perspective .. 28515
Yu Mitsuzumi (NTT Corporation, Japan; Kyoto University, Japan), Akisato Kimura (NTT Corporation, Japan), and Hisashi Kashima (Kyoto University, Japan)

Resurrecting Old Classes with New Data for Exemplar-Free Continual Learning 28525
Dipam Goswami (Computer Vision Center, Spain; Autonomous University of Barcelona, Spain), Albin Soutif-Cormerais (Computer Vision Center, Spain; Autonomous University of Barcelona, Spain), Yuyang Liu (University of Chinese Academy of Sciences, China; Shenyang Institute of Automation, Chinese Academy of Sciences, China), Sandesh Kamath (Computer Vision Center, Spain; Autonomous University of Barcelona, Spain), Bartłomiej Twardowski (Computer Vision Center, Spain; Autonomous University of Barcelona, Spain; IDEAS NCBR, Poland), and Joost van de Weijer (Computer Vision Center, Spain; Autonomous University of Barcelona, Spain)

Adversarially Robust Few-shot Learning via Parameter Co-distillation of Similarity and Class Concept Learners 28535
*Junhao Dong (Nanyang Technological University, Singapore and A*STAR, Singapore), Piotr Koniusz (CSIRO's Data61, Australia and Australian National University, Australia), Junxi Chen (Sun Yat-sen University, China), Xiaohua Xie (Sun Yat-sen University, China), and Yew-Soon Ong (Nanyang Technological University, Singapore and A*STAR, Singapore)*

Learning CNN on ViT: A Hybrid Model to Explicitly Class-specific Boundaries for Domain Adaptation 28545
Ba Hung Ngo (Chonnam National University), Nhat-Tuong Do-Tran (National Yang Ming Chiao Tung University), Tuan-Ngoc Nguyen (FPT Telecom), Hae-Gon Jeon (Gwangju Institute of Science and Technology), and Tae Jong Choi (Chonnam National University)

Efficient Stitchable Task Adaptation 28555
Haoyu He (Monash University), Zizheng Pan (Monash University), Jing Liu (Monash University), Jianfei Cai (Monash University), and Bohan Zhuang (Monash University)

Gradient-based Parameter Selection for Efficient Fine-Tuning 28566
Zhi Zhang (University of Amsterdam, Netherlands), Qizhe Zhang (Peking University, China), Zijun Gao (Shandong University, China), Renrui Zhang (The Chinese University of Hong Kong, China), Ekaterina Shutova (University of Amsterdam, Netherlands), Shiji Zhou (Tsinghua University), and Shanghang Zhang (Peking University, China)

ArGue: Attribute-Guided Prompt Tuning for Vision-Language Models 28578
Xinyu Tian (Australian National University, Australia), Shu Zou (Australian National University, Australia), Zhaoyuan Yang (GE Research, USA), and Jing Zhang (Australian National University, Australia)

Simple Semantic-Aided Few-Shot Learning 28588
Hai Zhang (Sichuan University, China), Junzhe Xu (Sichuan University, China; Alibaba Group, China), Shanlin Jiang (University of Texas at Dallas, USA), and Zhenan He (Sichuan University, China)

Long-Tail Class Incremental Learning via Independent Sub-prototype Construction 28598
Xi Wang (Xidian University, China), Xu Yang (Xidian University, China), Jie Yin (Xidian University, China), Kun Wei (Xidian University, China), and Cheng Deng (Xidian University, China)

Few-Shot Object Detection with Foundation Models 28608
Guangxing Han (Columbia University) and Ser-Nam Lim (University of Central Florida)

Stronger, Fewer, & Superior: Harnessing Vision Foundation Models for Domain Generalized Semantic Segmentation 28619
Zhixiang Wei (University of Science and Technology of China), Lin Chen (University of Science and Technology of China, Shanghai AI Laboratory), Yi Jin (University of Science and Technology of China), Xiaoxiao Ma (University of Science and Technology of China), Tianle Liu (University of Science and Technology of China), Pengyang Ling (University of Science and Technology of China, Shanghai AI Laboratory), Ben Wang (University of Science and Technology of China), Huaian Chen (University of Science and Technology of China), and Jinjin Zheng (University of Science and Technology of China)

Continual Forgetting for Pre-trained Vision Models 28631
Hongbo Zhao (CASIA, China), Bolin Ni (CASIA, China), Junsong Fan (HKISI_CAS), Yuxi Wang (HKISI_CAS), Yuntao Chen (HKISI_CAS), Gaofeng Meng (CASIA, China), and Zhaoxiang Zhang (CASIA, China)

AETTA: Label-Free Accuracy Estimation for Test-Time Adaptation 28643
Taeckyung Lee (KAIST), Sorn Chottananurak (KAIST), Taesik Gong (Nokia Bell Labs), and Sung-Ju Lee (KAIST)

Continual-MAE: Adaptive Distribution Masked Autoencoders for Continual Test-Time Adaptation 28653
Jiaming Liu (Peking University, China), Ran Xu (Peking University, China), Senqiao Yang (Peking University, China), Renrui Zhang (The Chinese University of Hong Kong), Qizhe Zhang (Peking University, China), Zehui Chen (University of Science and Technology of China), Yandong Guo (AI2Robotics, China), and Shanghang Zhang (Peking University, China)

LEAD: Exploring Logit Space Evolution for Model Selection 28664
Zixuan Hu (Peking University, Beijing, China), Xiaotong Li (Peking University, Beijing, China), Shixiang Tang (The Chinese University of Hong Kong, Hongkong, China), Jun Liu (Singapore University of Technology and Design, Singapore), Yichun Hu (Peking University, Beijing, China), and Ling-Yu Duan (Peking University, Beijing, China; Peng Cheng Laboratory, Shenzhen, China)

Instance-based Max-margin for Practical Few-shot Recognition 28674
Minghao Fu (Nanjing University, China) and Ke Zhu (Nanjing University, China)

Domain Gap Embeddings for Generative Dataset Augmentation 28684
Yinong Oliver Wang (Carnegie Mellon University, USA), Younjoon Chung (Carnegie Mellon University, USA), Chen Henry Wu (Carnegie Mellon University, USA), and Fernando De la Torre (Carnegie Mellon University, USA)

JoAPR: Cleaning the Lens of Prompt Learning for Vision-Language Models 28695
Yuncheng Guo (Fudan University, China) and Xiaodong Gu (Fudan University, China)

Generative Multi-modal Models are Good Class Incremental Learners 28706
Xusheng Cao (Nankai University, China), Haori Lu (Nankai University, China), Linlan Huang (Nankai University, China), Xialei Liu (Nankai University, China), and Ming-Ming Cheng (Nankai University, China)

Dual Memory Networks: A Versatile Adaptation Approach for Vision-Language Models 28718
Yabin Zhang (HKPolyU), Wenjie Zhu (HKPolyu), Hui Tang (HKUST), Zhiyuan Ma (PolyU), Kaiyang Zhou (HKBU), and Lei Zhang (HKPolyU)

UniPT: Universal Parallel Tuning for Transfer Learning with Efficient Parameter and Memory... 28729
Haiwen Diao (Dalian University of Technology, China), Bo Wan (Katholieke Universiteit Leuven, Belgium), Ying Zhang (Tencent WeChat, China), Xu Jia (Dalian University of Technology, China), Huchuan Lu (Dalian University of Technology, China), and Long Chen (The Hong Kong University of Science and Technology, China)

Federated Generalized Category Discovery 28741
Nan Pu (University of Trento, Italy), Wenjing Li (Hefei University of Technology, China; University of Leeds, UK), Xingyuan Ji (Xi'an Jiaotong University, China; Leiden University, The Netherlands), Yalan Qin (Shanghai University, China), Nicu Sebe (University of Trento, Italy), and Zhun Zhong (Hefei University of Technology, China; University of Nottingham, UK)

Learning from One Continuous Video Stream 28751
João Carreira (Google DeepMind), Michael King (Google DeepMind), Viorica Patraucean (Google DeepMind), Dilara Gokay (Google DeepMind), Catalin Ionescu (Google DeepMind), Yi Yang (Google DeepMind), Daniel Zoran (Google DeepMind), Joseph Heyward (Google DeepMind), Carl Doersch (Google DeepMind), Yusuf Aytar (Google DeepMind), Dima Damen (University of Bristol, Google DeepMind), and Andrew Zisserman (Google DeepMind, University of Oxford)

OrCo: Towards Better Generalization via Orthogonality and Contrast for Few-Shot Class-Incremental Learning 28762
Noor Ahmed (Max Planck Institute for Informatics), Anna Kukleva (Max Planck Institute for Informatics), and Bernt Schiele (Max Planck Institute for Informatics)

SDDGR: Stable Diffusion-based Deep Generative Replay for Class Incremental Object Detection 28772
Junsu Kim (Ulsan National Institute of Science and Technology, Korea), Hoseong Cho (Ulsan National Institute of Science and Technology, Korea), Jihyeon Kim (Ulsan National Institute of Science and Technology, Korea), Yihalem Yimolal Tiruneh (Ulsan National Institute of Science and Technology, Korea), and Seungryul Baek (Ulsan National Institute of Science and Technology, Korea)

Active Domain Adaptation with False Negative Prediction for Object Detection 28782
Yuzuru Nakamura (Panasonic Holdings Corporation, Japan), Yasunori Ishii (Panasonic Holdings Corporation, Japan), and Takayoshi Yamashita (Chubu University, Japan)

Stationary Representations: Optimally Approximating Compatibility and Implications for Improved Model Replacements 28793
Niccolò Biondi (University of Florence, Italy), Federico Pernici (University of Florence, Italy), Simone Ricci (University of Florence, Italy), and Alberto Del Bimbo (University of Florence, Italy)

Your Transferability Barrier is Fragile: Free-Lunch for Transferring the Non-Transferable Learning 28805
Ziming Hong (The University of Sydney, Australia), Li Shen (JD Explore Academy, China), and Tongliang Liu (The University of Sydney, Australia)

Transductive Zero-Shot and Few-Shot CLIP 28816
Ségolène Martin (Université Paris-Saclay, Inria, CentraleSupélec, France), Yunshi Huang (ETS Montréal, Canada), Fereshteh Shakeri (ETS Montréal, Canada), Jean-Christophe Pesquet (Université Paris-Saclay, Inria, CentraleSupélec, France), and Ismail Ben Ayed (ETS Montréal, Canada)

Task2Box: Box Embeddings for Modeling Asymmetric Task Relationships 28827
Rangel Daroya (University of Massachusetts Amherst), Aaron Sun (University of Massachusetts Amherst), and Subhransu Maji (University of Massachusetts Amherst)

Unbiased Faster R-CNN for Single-source Domain Generalized Object Detection …… 28838
Yajing Liu (Shenyang Institute of Automation, Chinese Academy of Sciences, China; Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), Shijun Zhou (Shenyang Institute of Automation, Chinese Academy of Sciences, China; Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, China; University), Xiyao Liu (Shenyang Institute of Automation, Chinese Academy of Sciences, China; Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, China), Chunhui Hao (Shenyang Institute of Automation, Chinese Academy of Sciences, China; Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, China), Baojie Fan (Nanjing University of Posts and Telecommunications, China), and Jiandong Tian (Shenyang Institute of Automation, Chinese Academy of Sciences, China; Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, China)

Author Index