

# **37th AAAI Conference on Artificial Intelligence (AAAI-23)**

## **Volume 3: AAAI Technical Tracks**

- Computer Vision III

Washington, DC, USA  
7-14 February 2023

Part 1 of 2

ISBN: 978-1-7138-7813-1

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2023) by Association for the Advancement of Artificial Intelligence  
All rights reserved.

Printed with permission by Curran Associates, Inc. (2023)

For permission requests, please contact Association for the Advancement of Artificial Intelligence  
at the address below.

Association for the Advancement of Artificial Intelligence  
2275 East Bayshore Road  
Suite 160  
Palo Alto, California 94303  
USA

Phone: 1-650-328-3123  
Fax: 1-650-321-4457

<https://aaai.org/Press/press.php>

**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  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

# TABLE OF CONTENTS

## PART 1

### AAAI TECHNICAL TRACK ON COMPUTER VISION III

|  |      |
|--|------|
| Ultra-High-Definition Low-Light Image Enhancement: A Benchmark and Transformer-Based Method .....  | 2654 |
| <i>Tao Wang, Kaihao Zhang, Tianrun Shen, Wenhan Luo, Bjorn Stenger, Tong Lu</i>                    |      |
| 3D Assembly Completion.....  | 2663 |
| <i>Weihao Wang, Rufeng Zhang, Mingyu You, Hongjun Zhou, Bin He</i>                                 |      |
| A Benchmark and Asymmetrical-Similarity Learning for Practical Image Copy Detection .....          | 2672 |
| <i>Wenhao Wang, Yifan Sun, Yi Yang</i>   |      |
| Revisiting Unsupervised Local Descriptor Learning.....   | 2680 |
| <i>Wufan Wang, Lei Zhang, Hua Huang</i>  |      |
| Crafting Monocular Cues and Velocity Guidance for Self-Supervised Multi-Frame Depth Learning ..... | 2689 |
| <i>Xiaofeng Wang, Zheng Zhu, Guan Huang, Xu Chi, Yun Ye, Ziwei Chen, Xingang Wang</i>              |      |
| Learning Continuous Depth Representation via Geometric Spatial Aggregator .....                    | 2698 |
| <i>Xiaohang Wang, Xuanhong Chen, Bingbing Ni, Zhengyan Tong, Hang Wang</i>                         |      |
| SSDA3D: Semi-supervised Domain Adaptation for 3D Object Detection from Point Cloud.....            | 2707 |
| <i>Yan Wang, Junbo Yin, Wei Li, Pascal Frossard, Ruigang Yang, Jianbing Shen</i>                   |      |
| High-Resolution GAN Inversion for Degraded Images in Large Diverse Datasets .....                  | 2716 |
| <i>Yanbo Wang, Chuming Lin, Donghao Luo, Ying Tai, Zhizhong Zhang, Yuan Xie</i>                    |      |
| GAN Prior Based Null-Space Learning for Consistent Super-resolution.....                           | 2724 |
| <i>Yinhuai Wang, Yujie Hu, Jiwen Yu, Jian Zhang</i>  |      |
| Contrastive Masked Autoencoders for Self-Supervised Video Hashing.....                             | 2733 |
| <i>Yuting Wang, Jinpeng Wang, Bin Chen, Ziyun Zeng, Shu-Tao Xia</i>                                |      |
| MicroAST: Towards Super-fast Ultra-Resolution Arbitrary Style Transfer .....                       | 2742 |
| <i>Zhizhong Wang, Lei Zhao, Zhiwen Zuo, Ailin Li, Haibo Chen, Wei Xing, Dongming Lu</i>            |      |
| Truncate-Split-Contrast: A Framework for Learning from Mislabeled Videos .....                     | 2751 |
| <i>Zixiao Wang, Junwu Weng, Chun Yuan, Jue Wang</i>  |      |
| Active Token Mixer.....  | 2759 |
| <i>Guoqiang Wei, Zhizheng Zhang, Cuiling Lan, Yan Lu, Zhibo Chen</i>                               |      |
| Exploring Non-target Knowledge for Improving Ensemble Universal Adversarial Attacks .....          | 2768 |
| <i>Juanjuan Weng, Zhiming Luo, Zhun Zhong, Dazhen Lin, Shaozi Li</i>                               |      |
| Towards Good Practices for Missing Modality Robust Action Recognition .....                        | 2776 |
| <i>Sangmin Woo, Sumin Lee, Yeonju Park, Muhammad Adi Nugroho, Changick Kim</i>                     |      |
| Reject Decoding via Language-Vision Models for Text-to-Image Synthesis.....                        | 2785 |
| <i>Fuxiang Wu, Liu Liu, Fusheng Hao, Fengxiang He, Lei Wang, Jun Cheng</i>                         |      |

|   |      |
|---|------|
| Transformation-Equivariant 3D Object Detection for Autonomous Driving .....                                     | 2795 |
| <i>Hai Wu, Chenglu Wen, Wei Li, Xin Li, Ruigang Yang, Cheng Wang</i>  |      |
| Super-efficient Echocardiography Video Segmentation via Proxy- and Kernel-Based Semi-supervised Learning.....   | 2803 |
| <i>Huisi Wu, Jingyin Lin, Wende Xie, Jing Qin</i>   |      |
| ACL-Net: Semi-supervised Polyp Segmentation via Affinity Contrastive Learning .....                             | 2812 |
| <i>Huisi Wu, Wende Xie, Jingyin Lin, Xinrong Guo</i>  |      |
| Bi-directional Feature Reconstruction Network for Fine-Grained Few-Shot Image Classification .....              | 2821 |
| <i>Jijie Wu, Dongliang Chang, Aneeshan Sain, Xiaoxu Li, Zhanyu Ma, Jie Cao, Jun Guo, Yi-Zhe Song</i>            |      |
| Preserving Structural Consistency in Arbitrary Artist and Artwork Style Transfer .....                          | 2830 |
| <i>Jingyu Wu, Lefan Hou, Zejian Li, Jun Liao, Li Liu, Lingyun Sun</i>   |      |
| End-to-End Zero-Shot HOI Detection via Vision and Language Knowledge Distillation .....                         | 2839 |
| <i>Mingrui Wu, Jiaxin Gu, Yunhang Shen, Mingbao Lin, Chao Chen, Xiaoshuai Sun</i>                               |      |
| Revisiting Classifier: Transferring Vision-Language Models for Video Recognition .....                          | 2847 |
| <i>Wenhao Wu, Zhun Sun, Wanli Ouyang</i>  |      |
| Scene Graph to Image Synthesis via Knowledge Consensus .....  | 2856 |
| <i>Yang Wu, Pengxu Wei, Liang Lin</i>   |      |
| Synthetic Data Can Also Teach: Synthesizing Effective Data for Unsupervised Visual Representation Learning..... | 2866 |
| <i>Yawen Wu, Zhepeng Wang, Dewen Zeng, Yiyu Shi, Jingtong Hu</i>  |      |
| Multi-Stream Representation Learning for Pedestrian Trajectory Prediction.....                                  | 2875 |
| <i>Yuxuan Wu, Le Wang, Sanping Zhou, Jinghai Duan, Gang Hua, Wei Tang</i>                                       |      |
| Pixel Is All You Need: Adversarial Trajectory-Ensemble Active Learning for Salient Object Detection .....       | 2883 |
| <i>Zhenyu Wu, Lin Wang, Wei Wang, Qing Xia, Chenglizhao Chen, Aimin Hao, Shuo Li</i>                            |      |
| Attention-Based Depth Distillation with 3D-Aware Positional Encoding for Monocular 3D Object Detection .....    | 2892 |
| <i>Zizhang Wu, Yunzhe Wu, Jian Pu, Xianzhi Li, Xiaoquan Wang</i>  |      |
| Skating-Mixer: Long-Term Sport Audio-Visual Modeling with MLPs.....   | 2901 |
| <i>Jingfei Xia, Mingchen Zhuge, Tiantian Geng, Shun Fan, Yuntai Wei, Zhenyu He, Feng Zheng</i>                  |      |
| SVFI: Spiking-Based Video Frame Interpolation for High-Speed Motion .....                                       | 2910 |
| <i>Lujie Xia, Jing Zhao, Ruiqin Xiong, Tiejun Huang</i>   |      |
| FEditNet: Few-Shot Editing of Latent Semantics in GAN Spaces.....   | 2919 |
| <i>Mengfei Xia, Yezhi Shu, Yuji Wang, Yu-Kun Lai, Qiang Li, Pengfei Wan, Zhongyuan Wang, Yong-Jin Liu</i>       |      |
| Toward Robust Diagnosis: A Contour Attention Preserving Adversarial Defense for COVID-19 Detection .....        | 2928 |
| <i>Kun Xiang, Xing Zhang, Jinwen She, Jinpeng Liu, Haohan Wang, Shiqi Deng, Shancheng Jiang</i>                 |      |

|  |      |
|--|------|
| Boosting Semi-Supervised Semantic Segmentation with Probabilistic Representations .....                                    | 2938 |
| <i>Haoyu Xie, Changqi Wang, Mingkai Zheng, Minjing Dong, Shan You, Chong Fu, Chang Xu</i>                                  |      |
| Less Is More Important: An Attention Module Guided by Probability Density Function for Convolutional Neural Networks ..... | 2947 |
| <i>Jingfen Xie, Jian Zhang</i>   |      |
| Mitigating Artifacts in Real-World Video Super-resolution Models .....   | 2956 |
| <i>Liangbin Xie, Xintao Wang, Shuwei Shi, Jinjin Gu, Chao Dong, Ying Shan</i>  |      |
| Just Noticeable Visual Redundancy Forecasting: A Deep Multimodal-Driven Approach.....                                      | 2965 |
| <i>Wuyuan Xie, Shukang Wang, Sukun Tian, Lirong Huang, Ye Liu, Miaohui Wang</i>  |      |
| Cross-Modal Contrastive Learning for Domain Adaptation in 3D Semantic Segmentation .....                                   | 2974 |
| <i>Bowei Xing, Xianghua Ying, Ruibin Wang, Jinfa Yang, Taiyan Chen</i>   |      |
| ROIFormer: Semantic-Aware Region of Interest Transformer for Efficient Self-Supervised Monocular Depth Estimation.....     | 2983 |
| <i>Daitao Xing, Jinglin Shen, Chiuman Ho, Anthony Tzes</i>   |      |
| LORE: Logical Location Regression Network for Table Structure Recognition .....  | 2992 |
| <i>Hangdi Xing, Feiyu Gao, Rujiao Long, Jiajun Bu, Qi Zheng, Liangcheng Li, Cong Yao, Zhi Yu</i>                           |      |
| Revisiting the Spatial and Temporal Modeling for Few-Shot Action Recognition .....   | 3001 |
| <i>Jiazheng Xing, Mengmeng Wang, Yong Liu, Boyu Mu</i>   |      |
| Unsupervised Multi-Exposure Image Fusion Breaking Exposure Limits via Contrastive Learning.....                            | 3010 |
| <i>Han Xu, Liang Haochen, Jiayi Ma</i>   |      |
| CasFusionNet: A Cascaded Network for Point Cloud Semantic Scene Completion by Dense Feature Fusion.....                    | 3018 |
| <i>Jinfeng Xu, Xianzhi Li, Yuan Tang, Qiao Yu, Yixue Hao, Long Hu, Min Chen</i>  |      |
| Learning a Generalized Gaze Estimator from Gaze-Consistent Feature .....   | 3027 |
| <i>Mingjie Xu, Haofei Wang, Feng Lu</i>  |      |
| Class Overwhelms: Mutual Conditional Blended-Target Domain Adaptation.....   | 3036 |
| <i>Pengcheng Xu, Boyu Wang, Charles Ling</i>   |      |
| Self Correspondence Distillation for End-to-End Weakly-Supervised Semantic Segmentation .....                              | 3045 |
| <i>Rongtao Xu, Changwei Wang, Jiayi Sun, Shibiao Xu, Weiliang Meng, Xiaopeng Zhang</i>                                     |      |
| Deep Parametric 3D Filters for Joint Video Denoising and Illumination Enhancement in Video Super Resolution .....          | 3054 |
| <i>Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, Jiaya Jia</i>   |      |
| Inter-image Contrastive Consistency for Multi-Person Pose Estimation .....   | 3063 |
| <i>Xixia Xu, Yingguo Gao, Xingjia Pan, Ke Yan, Xiaoyu Chen, Qi Zou</i>   |      |
| DeMT: Deformable Mixer Transformer for Multi-Task Learning of Dense Prediction .....                                       | 3072 |
| <i>Yangyang Xu, Yibo Yang, Lefei Zhang</i>   |      |
| VLtinT: Visual-Linguistic Transformer-in-Transformer for Coherent Video Paragraph Captioning .....                         | 3081 |
| <i>Kashu Yamazaki, Khoa Vo, Quang Sang Truong, Bhiksha Raj, Ngan Le</i>  |      |
| Rethinking Disparity: A Depth Range Free Multi-View Stereo Based on Disparity .....  | 3091 |
| <i>Qingsong Yan, Qiang Wang, Kaiyong Zhao, Bo Li, Xiaowen Chu, Fei Deng</i>  |      |

|   |      |
|---|------|
| Video-Text Pre-training with Learned Regions for Retrieval .....  | 3100 |
| <i>Rui Yan, Mike Zheng Shou, Yixiao Ge, Jinpeng Wang, Xudong Lin, Guanyu Cai, Jinhui Tang</i>                       |      |
| DesNet: Decomposed Scale-Consistent Network for Unsupervised Depth Completion.....                                  | 3109 |
| <i>Zhiqiang Yan, Kun Wang, Xiang Li, Zhenyu Zhang, Jun Li, Jian Yang</i>  |      |
| Self-Supervised Video Representation Learning via Latent Time Navigation .....                                      | 3118 |
| <i>Di Yang, Yaohui Wang, Quan Kong, Antitza Dantcheva, Lorenzo Garattoni, Gianpiero Francesca, François Brémond</i> |      |
| One-Shot Replay: Boosting Incremental Object Detection via Retrospecting One Object.....                            | 3127 |
| <i>Dongbao Yang, Yu Zhou, Xiaopeng Hong, Aoting Zhang, Weiping Wang</i>   |      |
| Video Event Extraction via Tracking Visual States of Arguments .....  | 3136 |
| <i>Guang Yang, Manling Li, Jiajie Zhang, Xudong Lin, Heng Ji, Shih-Fu Chang</i>                                     |      |
| CoMAE: Single Model Hybrid Pre-training on Small-Scale RGB-D Datasets.....  | 3145 |
| <i>Jiange Yang, Sheng Guo, Gangshan Wu, Limin Wang</i>  |      |
| Self-Asymmetric Invertible Network for Compression-Aware Image Rescaling.....                                       | 3155 |
| <i>Jinhai Yang, Mengxi Guo, Shijie Zhao, Junlin Li, Li Zhang</i>  |      |
| Stop-Gradient Softmax Loss for Deep Metric Learning.....  | 3164 |
| <i>Lu Yang, Peng Wang, Yanning Zhang</i>  |      |
| Local Path Integration for Attribution .....  | 3173 |
| <i>Peiyu Yang, Naveed Akhtar, Zeyi Wen, Ajmal Mian</i>  |      |
| Spatiotemporal Deformation Perception for Fisheye Video Rectification.....  | 3181 |
| <i>Shangrong Yang, Chunyu Lin, Kang Liao, Yao Zhao</i>  |      |
| Contrastive Multi-Task Dense Prediction.....  | 3190 |
| <i>Siwei Yang, Hanrong Ye, Dan Xu</i>   |      |
| AutoStegaFont: Synthesizing Vector Fonts for Hiding Information in Documents .....                                  | 3198 |
| <i>Xi Yang, Jie Zhang, Han Fang, Chang Liu, Zehua Ma, Weiming Zhang, Nenghai Yu</i>                                 |      |
| Towards Global Video Scene Segmentation with Context-Aware Transformer .....  | 3206 |
| <i>Yang Yang, Yurui Huang, Weili Guo, Baohua Xu, Dingyin Xia</i>  |      |
| Low-Light Image Enhancement Network Based on Multi-Scale Feature Complementation.....                               | 3214 |
| <i>Yong Yang, Wenzhi Xu, Shuying Huang, Weiguo Wan</i>  |      |
| Semantics-Aware Dynamic Localization and Refinement for Referring Image Segmentation .....                          | 3222 |
| <i>Zhao Yang, Jiaqi Wang, Yansong Tang, Kai Chen, Hengshuang Zhao, Philip H.S. Torr</i>                             |      |
| LidarMultiNet: Towards a Unified Multi-Task Network for LiDAR Perception .....                                      | 3231 |
| <i>Dongqiangzi Ye, Zixiang Zhou, Weijia Chen, Yufei Xie, Yu Wang, Panqu Wang, Hassan Foroosh</i>                    |      |
| DPTText-DETR: Towards Better Scene Text Detection with Dynamic Points in Transformer .....                          | 3241 |
| <i>Maoyuan Ye, Jing Zhang, Shanshan Zhao, Juhua Liu, Bo Du, Dacheng Tao</i>   |      |
| Learning Second-Order Attentive Context for Efficient Correspondence Pruning .....                                  | 3250 |
| <i>Xinyi Ye, Weyue Zhao, Hao Lu, Zhiguo Cao</i>   |      |

Infusing Definiteness into Randomness: Rethinking Composition Styles for Deep Image Matting ..... 3259  
*Zixuan Ye, Yutong Dai, Chaoyi Hong, Zhiguo Cao, Hao Lu*

## PART 2

Can We Find Strong Lottery Tickets in Generative Models? ..... 3267  
*Sangyeop Yeo, Yoojin Jang, Jy-yong Sohn, Dongyoon Han, Jaejun Yoo*

Class-Independent Regularization for Learning with Noisy Labels ..... 3276  
*Rumeng Yi, Dayan Guan, Yaping Huang, Shijian Lu*

Unbiased Heterogeneous Scene Graph Generation with Relation-Aware Message Passing Neural Network ..... 3285  
*Kanghoon Yoon, Kibum Kim, Jinyoung Moon, Chanyoung Park*

Lifelong Person Re-identification via Knowledge Refreshing and Consolidation ..... 3295  
*Chunlin Yu, Ye Shi, Zimo Liu, Shenghua Gao, Jingya Wang*

Generalizing Multiple Object Tracking to Unseen Domains by Introducing Natural Language Representation ..... 3304  
*En Yu, Songtao Liu, Zhuoling Li, Jinrong Yang, Zeming Li, Shoudong Han, Wenbing Tao*

Rethinking Rotation Invariance with Point Cloud Registration ..... 3313  
*Jianhui Yu, Chaoyi Zhang, Weidong Cai*

Frame-Level Label Refinement for Skeleton-Based Weakly-Supervised Action Recognition ..... 3322  
*Qing Yu, Kent Fujiwara*

Recurrent Structure Attention Guidance for Depth Super-resolution ..... 3331  
*Jiayi Yuan, Haobo Jiang, Xiang Li, Jianjun Qian, Jun Li, Jian Yang*

Structure Flow-Guided Network for Real Depth Super-resolution ..... 3340  
*Jiayi Yuan, Haobo Jiang, Xiang Li, Jianjun Qian, Jun Li, Jian Yang*

Pseudo Label-Guided Model Inversion Attack via Conditional Generative Adversarial Network ..... 3349  
*Xiaojian Yuan, Kejiang Chen, Jie Zhang, Weiming Zhang, Nenghai Yu, Yang Zhang*

Cyclically Disentangled Feature Translation for Face Anti-spoofing ..... 3358  
*Haixiao Yue, Keyao Wang, Guosheng Zhang, Haocheng Feng, Junyu Han, Errui Ding, Jingdong Wang*

FlowFace: Semantic Flow-Guided Shape-Aware Face Swapping ..... 3367  
*Hao Zeng, Wei Zhang, Changjie Fan, Tangjie Lv, Suzhen Wang, Zhimeng Zhang, Bowen Ma, Lincheng Li, Yu Ding, Xin Yu*

Multi-Modal Knowledge Hypergraph for Diverse Image Retrieval ..... 3376  
*Yawen Zeng, Qin Jin, Tengfei Bao, Wenfeng Li*

Learnable Blur Kernel for Single-Image Defocus Deblurring in the Wild ..... 3384  
*Jucai Zhai, Pengcheng Zeng, Chihao Ma, Jie Chen, Yong Zhao*

Darwinian Model Upgrades: Model Evolving with Selective Compatibility ..... 3393  
*Binjie Zhang, Shupeng Su, Yixiao Ge, Xuyuan Xu, Yexin Wang, Chun Yuan, Mike Zheng Shou, Ying Shan*

|   |      |
|---|------|
| Mx2M: Masked Cross-Modality Modeling in Domain Adaptation for 3D Semantic Segmentation.....   | 3401 |
| <i>Boxiang Zhang, Zunran Wang, Yonggen Ling, Yuanyuan Guan, Shenghao Zhang, Wenhui Li</i>   |      |
| Few-Shot 3D Point Cloud Semantic Segmentation via Stratified Class-Specific Attention Based<br>Transformer Network .....                | 3410 |
| <i>Canyu Zhang, Zhenyao Wu, Xinyi Wu, Ziyu Zhao, Song Wang</i>  |      |
| PaRot: Patch-Wise Rotation-Invariant Network via Feature Disentanglement and Pose Restoration.....                                      | 3418 |
| <i>Dingxin Zhang, Jianhui Yu, Chaoyi Zhang, Weidong Cai</i>   |      |
| Hierarchical Consistent Contrastive Learning for Skeleton-Based Action Recognition with<br>Growing Augmentations.....                   | 3427 |
| <i>Jiahang Zhang, Lilang Lin, Jiaying Liu</i>   |      |
| ImageNet Pre-training Also Transfers Non-robustness .....   | 3436 |
| <i>Jiaming Zhang, Jitao Sang, Qi Yi, Yunfan Yang, Huiwen Dong, Jian Yu</i>  |      |
| Language-Assisted 3D Feature Learning for Semantic Scene Understanding .....  | 3445 |
| <i>Junbo Zhang, Guofan Fan, Guanghan Wang, Zhengyuan Su, Kaisheng Ma, Li Yi</i>   |      |
| IKOL: Inverse Kinematics Optimization Layer for 3D Human Pose and Shape Estimation via<br>Gauss-Newton Differentiation.....             | 3454 |
| <i>Juze Zhang, Ye Shi, Yuexin Ma, Lan Xu, Jingyi Yu, Jingya Wang</i>  |      |
| Mind the Gap: Polishing Pseudo Labels for Accurate Semi-supervised Object Detection .....   | 3463 |
| <i>Lei Zhang, Yuxuan Sun, Wei Wei</i>   |      |
| ConvMatch: Rethinking Network Design for Two-View Correspondence Learning .....   | 3472 |
| <i>Shihua Zhang, Jiayi Ma</i>   |      |
| Cross-View Geo-Localization via Learning Disentangled Geometric Layout Correspondence .....   | 3480 |
| <i>Xiaohan Zhang, Xingyu Li, Waqas Sultani, Yi Zhou, Safwan Wshah</i>   |      |
| Video Compression Artifact Reduction by Fusing Motion Compensation and Global Context in a<br>Swin-CNN Based Parallel Architecture..... | 3489 |
| <i>Xinjian Zhang, Su Yang, Wuyang Luo, Longwen Gao, Weishan Zhang</i>   |      |
| MRCN: A Novel Modality Restitution and Compensation Network for Visible-Infrared Person Re-<br>identification.....                      | 3498 |
| <i>Yukang Zhang, Yan Yan, Jie Li, Hanzi Wang</i>  |      |
| A Simple Baseline for Multi-Camera 3D Object Detection .....  | 3507 |
| <i>Yunpeng Zhang, Wenzhao Zheng, Zheng Zhu, Guan Huang, Jiwen Lu, Jie Zhou</i>  |      |
| Positional Label for Self-Supervised Vision Transformer .....   | 3516 |
| <i>Zheming Zhang, Xun Gong</i>  |      |
| Cross-Category Highlight Detection via Feature Decomposition and Modality Alignment .....   | 3525 |
| <i>Zhenduo Zhang</i>  |      |
| TrEP: Transformer-Based Evidential Prediction for Pedestrian Intention with Uncertainty .....   | 3534 |
| <i>Zhengming Zhang, Renran Tian, Zhengming Ding</i>   |      |
| DINet: Deformation Inpainting Network for Realistic Face Visually Dubbing on High Resolution<br>Video .....                             | 3543 |
| <i>Zhimeng Zhang, Zhipeng Hu, Wenjin Deng, Changjie Fan, Tangjie Lv, Yu Ding</i>  |      |



|  |      |
|--|------|
| ShiftDDPMs: Exploring Conditional Diffusion Models by Shifting Diffusion Trajectories.....   | 3552 |
| <i>Zijian Zhang, Zhou Zhao, Jun Yu, Qi Tian</i>  |      |
| Combating Unknown Bias with Effective Bias-Conflicting Scoring and Gradient Alignment.....   | 3561 |
| <i>Bowen Zhao, Chen Chen, Qian-Wei Wang, Anfeng He, Shu-Tao Xia</i>  |      |
| RLogist: Fast Observation Strategy on Whole-Slide Images with Deep Reinforcement Learning .....  | 3570 |
| <i>Boxuan Zhao, Jun Zhang, Deheng Ye, Jian Cao, Xiao Han, Qiang Fu, Wei Yang</i>   |      |
| Learning to Super-resolve Dynamic Scenes for Neuromorphic Spike Camera.....  | 3579 |
| <i>Jing Zhao, Ruiqin Xiong, Jian Zhang, Rui Zhao, Hangfan Liu, Tiejun Huang</i>  |      |
| TinyNeRF: Towards 100 x Compression of Voxel Radiance Fields.....  | 3588 |
| <i>Tianli Zhao, Jiayuan Chen, Cong Leng, Jian Cheng</i>  |      |
| BEST: BERT Pre-training for Sign Language Recognition with Coupling Tokenization .....   | 3597 |
| <i>Weichao Zhao, Hezhen Hu, Wengang Zhou, Jiaxin Shi, Houqiang Li</i>  |      |
| MuGT: Multi-Task Graph-Transformer with Task-Aware Knowledge Injection and Domain<br>Knowledge-Driven Pooling for Whole Slide Image Analysis ..... | 3606 |
| <i>Wei Qin Zhao, Shujun Wang, Maximus Yeung, Tianye Niu, Lequan Yu</i>   |      |
| Grouped Knowledge Distillation for Deep Face Recognition.....  | 3615 |
| <i>Weisong Zhao, Xiangyu Zhu, Kaiwen Guo, Xiao-Yu Zhang, Zhen Lei</i>  |      |
| Style-Content Metric Learning for Multidomain Remote Sensing Object Recognition.....   | 3624 |
| <i>Wenda Zhao, Ruikai Yang, Yu Liu, You He</i>   |      |
| Occupancy Planes for Single-View RGB-D Human Reconstruction .....  | 3633 |
| <i>Xiaoming Zhao, Yuan-Ting Hu, Zhongzheng Ren, Alexander G. Schwing</i>   |      |
| Deep Equilibrium Models for Snapshot Compressive Imaging .....   | 3642 |
| <i>Yaping Zhao, Siming Zheng, Xin Yuan</i>   |      |
| Unsupervised Deep Video Denoising with Untrained Network .....   | 3651 |
| <i>Huan Zheng, Tongyao Pang, Hui Ji</i>  |      |
| Attack Can Benefit: An Adversarial Approach to Recognizing Facial Expressions under Noisy<br>Annotations .....                                     | 3660 |
| <i>Jiawen Zheng, Bo Li, Shengchuan Zhang, Shuang Wu, Liujuan Cao, Shouhong Ding</i>  |      |
| Phrase-Level Temporal Relationship Mining for Temporal Sentence Localization .....   | 3669 |
| <i>Minghang Zheng, Sizhe Li, Qingchao Chen, Yuxin Peng, Yang Liu</i>   |      |
| Learning Semantic Degradation-Aware Guidance for Recognition-Driven Unsupervised Low-Light<br>Image Enhancement .....                              | 3678 |
| <i>Naishan Zheng, Jie Huang, Man Zhou, Zizheng Yang, Qi Zhu, Feng Zhao</i>   |      |
| Memory-Aided Contrastive Consensus Learning for Co-salient Object Detection.....   | 3687 |
| <i>Peng Zheng, Jie Qin, Shuo Wang, Tian-Zhu Xiang, Huan Xiong</i>  |      |
| MaskBooster: End-to-End Self-Training for Sparsely Supervised Instance Segmentation.....   | 3696 |
| <i>Shida Zheng, Chenshu Chen, Xi Yang, Wenming Tan</i>   |      |
| RSPT: Reconstruct Surroundings and Predict Trajectory for Generalizable Active Object Tracking .....   | 3705 |
| <i>Fangwei Zhong, Xiao Bi, Yudi Zhang, Wei Zhang, Yizhou Wang</i>  |      |

|  |      |
|--|------|
| STOA-VLP: Spatial-Temporal Modeling of Object and Action for Video-Language Pre-training.....                    | 3715 |
| <i>Weihong Zhong, Mao Zheng, Duyu Tang, Xuan Luo, Heng Gong, Xiaocheng Feng, Bing Qin</i>                        |      |
| Refined Semantic Enhancement towards Frequency Diffusion for Video Captioning.....                               | 3724 |
| <i>Xian Zhong, Zipeng Li, Shuqin Chen, Kui Jiang, Chen Chen, Mang Ye</i>   |      |
| Aesthetically Relevant Image Captioning .....  | 3733 |
| <i>Zhipeng Zhong, Fei Zhou, Guoping Qiu</i>  |      |
| Polarization-Aware Low-Light Image Enhancement .....   | 3742 |
| <i>Chu Zhou, Minggui Teng, Youwei Lyu, Si Li, Chao Xu, Boxin Shi</i>   |      |
| Progressive Bayesian Inference for Scribble-Supervised Semantic Segmentation .....                               | 3751 |
| <i>Chuanwei Zhou, Chunyan Xu, Zhen Cui</i>   |      |
| Exploratory Inference Learning for Scribble Supervised Semantic Segmentation .....                               | 3760 |
| <i>Chuanwei Zhou, Zhen Cui, Chunyan Xu, Cao Han, Jian Yang</i>   |      |
| Dual Memory Units with Uncertainty Regulation for Weakly Supervised Video Anomaly Detection .....                | 3769 |
| <i>Hang Zhou, Junqing Yu, Wei Yang</i>   |      |
| Unsupervised Hierarchical Domain Adaptation for Adverse Weather Optical Flow.....                                | 3778 |
| <i>Hanyu Zhou, Yi Chang, Gang Chen, Luxin Yan</i>  |      |
| PASS: Patch Automatic Skip Scheme for Efficient Real-Time Video Perception on Edge Devices.....                  | 3787 |
| <i>Qihua Zhou, Song Guo, Jun Pan, Jiacheng Liang, Zhenda Xu, Jingren Zhou</i>                                    |      |
| Robust Feature Rectification of Pretrained Vision Models for Object Recognition.....                             | 3796 |
| <i>Shengchao Zhou, Gaofeng Meng, Zhaoxiang Zhang, Richard Yi Da Xu, Shiming Xiang</i>                            |      |
| Video Object of Interest Segmentation.....   | 3805 |
| <i>Siyuan Zhou, Chunru Zhan, Biao Wang, Tiezheng Ge, Yuning Jiang, Li Niu</i>                                    |      |
| Tree-Structured Trajectory Encoding for Vision-and-Language Navigation.....                                      | 3814 |
| <i>Xinzhe Zhou, Yadong Mu</i>  |      |
| Self-Supervised Action Representation Learning from Partial Spatio-Temporal Skeleton Sequences.....              | 3825 |
| <i>Yujie Zhou, Haodong Duan, Anyi Rao, Bing Su, Jiaqi Wang</i>   |      |
| Debiased Fine-Tuning for Vision-Language Models by Prompt Regularization.....                                    | 3834 |
| <i>Beier Zhu, Yulei Niu, Saeil Lee, Minhoe Hur, Hanwang Zhang</i>  |      |
| Improving Scene Text Image Super-resolution via Dual Prior Modulation Network.....                               | 3843 |
| <i>Shipeng Zhu, Zuoyan Zhao, Pengfei Fang, Hui Xue</i>   |      |
| SRoUDA: Meta Self-Training for Robust Unsupervised Domain Adaptation .....                                       | 3852 |
| <i>Wanqing Zhu, Jia-Li Yin, Bo-Hao Chen, Ximeng Liu</i>  |      |
| Gradient-Based Graph Attention for Scene Text Image Super-resolution .....                                       | 3861 |
| <i>Xiangyuan Zhu, Kehua Guo, Hui Fang, Rui Ding, Zheng Wu, Gerald Schaefer</i>                                   |      |
| RGBD1K: A Large-Scale Dataset and Benchmark for RGB-D Object Tracking.....                                       | 3870 |
| <i>Xue-Feng Zhu, Tianyang Xu, Zhangyong Tang, Zucheng Wu, Haodong Liu, Xiao Yang, Xiao-Jun Wu, Josef Kittler</i> |      |
| Learn More for Food Recognition via Progressive Self-Distillation.....   | 3879 |
| <i>Yaohui Zhu, Linhu Liu, Jiang Tian</i>   |      |

Generative Image Inpainting with Segmentation Confusion Adversarial Training and Contrastive Learning ..... 3888  
*Zhiwen Zuo, Lei Zhao, Ailin Li, Zhizhong Wang, Zhanjie Zhang, Jiafu Chen, Wei Xing, Dongming Lu*

**Author Index**