

**40th AAAI Conference on Artificial Intelligence (AAAI-26), 38th Conference on Innovative Applications of Artificial Intelligence (IAAI-26), and 16th Symposium on Educational Advances in Artificial Intelligence (EAAI-26)**

Volume 12: AAAI Technical Tracks

- Computer Vision IX

Singapore  
20-27 January 2026

ISBN: 979-8-3313-3386-7

**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© (2026) by Association for the Advancement of Artificial Intelligence  
All rights reserved.

Printed with permission by Curran Associates, Inc. (2026)

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

## AAAI TECHNICAL TRACK ON COMPUTER VISION IX

|  |      |
|--|------|
| CoGrad3D: Spatially-Coupled Timestep Optimization with Orthogonal Gradient Fusion for 3D Generation .....                                    | 9539 |
| <i>Haoyang Tong, Hongbo Wang, Jin Liu, Qi Wang, Jie Cao, Ran He</i>  |      |
| Towards Accurate 3D Object Detection in Adverse Weather by Leveraging 4D Radar for LiDAR Geometry Enhancement .....                          | 9548 |
| <i>Tianxu Tong, Xinrun Liu, Hongmin Liu, Bin Fan</i>   |      |
| CADiff: Context-Aware Diffusion for Controllable Anomaly Generation in Anomaly Detection.....  | 9557 |
| <i>Xuan Tong, Yuxuan Lin, Junxiong Lin, Xinji Mai, Haoran Wang, Zeng Tao, Yang Yao, Ruofan Wang, Wenqiang Zhang</i>                          |      |
| FedSDA: Federated Stain Distribution Alignment for Non-IID Histopathological Image Classification .....                                      | 9566 |
| <i>Cheng-Chang Tsai, Kai-Wen Cheng, Chun-Shien Lu</i>  |      |
| Mitigating Low-Quality Reasoning in MLLMs: Self-Driven Refined Multimodal CoT with Selective Thinking and Step-Wise Visual Enhancement ..... | 9576 |
| <i>Chongjun Tu, Peng Ye, Dongzhan Zhou, Tao Chen, Wanli Ouyang</i>   |      |
| Mass Concept Erasure in Diffusion Models with Concept Hierarchy .....  | 9585 |
| <i>Jiahang Tu, Ye Li, Yiming Wu, Hanbin Zhao, Chao Zhang, Hui Qian</i>   |      |
| DiLO: Disentangled Latent Optimization for Learning Shape and Deformation in Grouped Deforming 3D Objects .....                              | 9594 |
| <i>Mostofa Rafid Uddin, Jana Armouti, Umong Sain, Md Asib Rahman, Xingjian Li, Min Xu</i>  |      |
| NURBGen: High-Fidelity Text-To-CAD Generation Through LLM-Driven NURBS Modeling.....   | 9603 |
| <i>Muhammad Usama, Mohammad Sadil Khan, Didier Stricker, Muhammad Zeshan Afzal</i>   |      |
| Guideline-Consistent Segmentation Via Multi-Agent Refinement .....   | 9612 |
| <i>Vanshika Vats, Ashwani Rathee, James Davis</i>  |      |
| Certified but Fooled! Breaking Certified Defenses with Ghost Certificates .....  | 9621 |
| <i>Quoc Viet Vo, Tashreque Mohammed Haq, Paul Montague, Tamas Abraham, Ehsan Abbasnejad, Damith C. Ranasinghe</i>                            |      |
| GUSLO: General and Unified Structured Light Optimization .....   | 9630 |
| <i>Tinglei Wan, Zhongjie Wang, Tonghua Su</i>  |      |
| Tuning Medical Foundation Models for Inner Ear Temporal CT Analysis with Plug-And-Play Domain Knowledge Aggregator .....                     | 9639 |
| <i>Weixun Wan, Xinyang Jiang, Zilong Wang, Bei Li, Cairong Zhao</i>  |      |
| Diffusion-Based Personalized Pathology Disentanglement for Impaired Gait Analysis .....  | 9648 |
| <i>Xiaoyue Wan, Xu Zhao</i>  |      |

|   |      |
|---|------|
| Perceptual Quality Assessment of 3D Gaussian Splatting: A Subjective Dataset and Prediction Metric .....  | 9657 |
| <i>Zhaolin Wan, Yining Diao, Jingqi Xu, Hao Wang, Zhiyang Li, Xiaopeng Fan, Wangmeng Zuo, Debin Zhao</i>  |      |
| Biologically-Inspired Evolutionary Domain Symbiosis for Few-Shot and Zero-Shot Point Cloud Semantic Segmentation .....  | 9666 |
| <i>Changshuo Wang, Zhijian Hu, Xiang Fang, Zai Yang Yu, Yibin Wu, Mingkun Xu, Yusong Wang, Xingyu Gao, Prayag Tiwari</i>  |      |
| AEDR: Training-Free AI-Generated Image Attribution Via Autoencoder Double-Reconstruction .....  | 9675 |
| <i>Chao Wang, Zijin Yang, Yaofei Wang, Weiming Zhang, Kejiang Chen</i>  |      |
| SGS-3D: High-Fidelity 3D Instance Segmentation Via Reliable Semantic Mask Splitting and Growing.....  | 9684 |
| <i>Chaolei Wang, Yang Luo, Jing Du, Siyu Chen, Yiping Chen, Ting Han</i>  |      |
| S <sup>2</sup> Flow: Towards Fast and Authentic Training-Free High-Resolution Video Generation.....   | 9693 |
| <i>Chaoqun Wang, Shaobo Min, Xu Yang</i>  |      |
| Axis-Aligned Document Dewarping .....   | 9702 |
| <i>Chaoyun Wang, I-Chao Shen, Takeo Igarashi, Caigui Jiang</i>  |      |
| Self-NPO: Data-Free Diffusion Model Enhancement Via Truncated Diffusion Fine-Tuning .....   | 9711 |
| <i>Fu-Yun Wang, Keqiang Sun, Yao Teng, Xihui Liu, Jiale Yuan, Jiaming Song, Hongsheng Li</i>  |      |
| 3One2: One-Step Regression Plus One-Step Diffusion for One-Hot Modulation in Dual-Path Video Snapshot Compressive Imaging .....   | 9720 |
| <i>Ge Wang, Xing Liu, Xin Yuan</i>  |      |
| UniMo: Unified Motion Generation and Understanding with Chain of Thought.....   | 9729 |
| <i>Guocun Wang, Kenkun Liu, Jing Lin, Guorui Song, Jian Li, Xiaoguang Han</i>   |      |
| Affordance-R1: Reinforcement Learning for Generalizable Affordance Reasoning in Multimodal Large Language Models .....  | 9738 |
| <i>Hanqing Wang, Shaoyang Wang, Yiming Zhong, Zemin Yang, Jiamin Wang, Zhiqing Cui, Jiahao Yuan, Yifan Han, Mingyu Liu, Yuexin Ma</i>   |      |
| VIR-Bench: Evaluating Geospatial and Temporal Understanding of MLLMs Via Travel Video Itinerary Reconstruction .....  | 9747 |
| <i>Hao Wang, Eiki Murata, Lingfang Zhang, Ayako Sato, So Fukuda, Ziqi Yin, Wentao Hu, Keisuke Nakao, Yusuke Nakamura, Sebastian Zwirner, Yi-Chia Chen, Hiroyuki Otomo, Hiroki Ouchi, Daisuke Kawahara</i> |      |
| SOMA: Feature Gradient Enhanced Affine-Flow Matching for SAR-Optical Registration .....   | 9757 |
| <i>Haodong Wang, Tao Zhuo, Xiuwei Zhang, Hanlin Yin, Wencong Wu, Yanning Zhang</i>  |      |
| READ: Real-Time and Efficient Asynchronous Diffusion for Audio-Driven Talking Head Generation .....   | 9766 |
| <i>Haotian Wang, Yuzhe Weng, Jun Du, Haoran Xu, Xiaoyan Wu, Shan He, Bing Yin, Cong Liu, Jianqing Gao, Qingfeng Liu</i>   |      |
| JoDiffusion: Jointly Diffusing Image with Pixel-Level Annotations for Semantic Segmentation Promotion.....  | 9775 |
| <i>Haoyu Wang, Lei Zhang, Wenrui Liu, Dengyang Jiang, Wei Wei, Chen Ding</i>  |      |

|  |      |
|--|------|
| RTGaze: Real-Time 3D-Aware Gaze Redirection from a Single Image .....  | 9784 |
| <i>Hengfei Wang, Zhongqun Zhang, Yihua Cheng, Hyung Jin Chang</i>  |      |
| Temporal-Consistent Video Restoration with Pre-Trained Diffusion Models .....  | 9793 |
| <i>Hengkang Wang, Yang Liu, Huidong Liu, Chien-Chih Wang, Yanhui Guo, Hongdong Li, Bryan Wang, Ju Sun</i>  |      |
| HarmoQ: Harmonized Post-Training Quantization for High-Fidelity Image Super-Resolution .....   | 9802 |
| <i>Hongjun Wang, Jiyuan Chen, Xuan Song, Yinqiang Zheng</i>  |      |
| FaceShield: Explainable Face Anti-Spoofing with Multimodal Large Language Models .....   | 9811 |
| <i>Hongyang Wang, Yichen Shi, Zhuofu Tao, Yuhao Gao, Liepiao Zhang, Xun Lin, Jun Feng, Xiaochen Yuan, Zitong Yu, Xiaochun Cao</i>                    |      |
| MoEA-Net: Modality-Incremental Expert Aggregation Network for Retinal Prognostic Prediction.....   | 9820 |
| <i>Hua Wang, Xiaodan Zhang, Yanzhao Shi, Chengxin Zheng, Wanyu Zhang, Zhen Wang, Jianing Wang, Xiaobing Yu</i>                                       |      |
| Multi-Window Gabor Transform Network for Ground Penetrating Radar B-Scan Image Reconstruction.....   | 9829 |
| <i>Huabin Wang, Yu Yang, Xinran Zhong, Zilong Ling</i>   |      |
| Towards Zero-Shot Diabetic Retinopathy Grading: Learning Generalized Knowledge Via Prompt-Driven Matching and Emulating .....                        | 9838 |
| <i>Huan Wang, Haoran Li, Yuxin Lin, Huaming Chen, Jun Yan, Lijuan Wang, Jiahua Shi, Qihao Xu, Yongting Hu, Yong Xu, Jun Shen</i>                     |      |
| Semantic Feature Purification for Adversarially-Aware RGB-T Tracking .....   | 9847 |
| <i>Jiahao Wang, Fang Liu, Hao Wang, Shuo Li, Xinyi Wang, Puhua Chen</i>  |      |
| HTTrack: Learning to Perceive Targets Via Historical Trajectories in Satellite Video Tracking .....  | 9856 |
| <i>Jiahao Wang, Fang Liu, Licheng Jiao, Hao Wang, Shuo Li, Xinyi Wang, Lingling Li, Puhua Chen, Xu Liu</i>   |      |
| Griffin: Aerial-Ground Cooperative Detection and Tracking Dataset and Benchmark .....  | 9867 |
| <i>Jiahao Wang, Xiangyu Cao, Jiaru Zhong, Yuner Zhang, Zeyu Han, Haibao Yu, Chuang Zhang, Lei He, Shaobing Xu, Jianqiang Wang</i>                    |      |
| SparseCoop: Cooperative Perception with Kinematic-Grounded Queries .....   | 9876 |
| <i>Jiahao Wang, Zhongwei Jiang, Wenchao Sun, Jiaru Zhong, Haibao Yu, Yuner Zhang, Chenyang Lu, Chuang Zhang, Lei He, Shaobing Xu, Jianqiang Wang</i> |      |
| EPSegFZ: Efficient Point Cloud Semantic Segmentation for Few- And Zero-Shot Scenarios with Language Guidance .....                                   | 9885 |
| <i>Jiahui Wang, Haiyue Zhu, Haoren Guo, Abdullah Al Mamun, Cheng Xiang, Tong Heng Lee</i>  |      |
| Monte Carlo Diffusion for Generalizable Learning-Based RANSAC .....  | 9894 |
| <i>Jiale Wang, Chen Zhao, Wei Ke, Tong Zhang</i>   |      |
| Invisible Triggers, Visible Threats! Road-Style Adversarial Creation Attack for Visual 3D Detection in Autonomous Driving .....                      | 9903 |
| <i>Jian Wang, Lijun He, Yixing Yong, Haixia Bi, Fan Li</i>   |      |
| SpaceVLLM: Endowing Multimodal Large Language Model with Spatio-Temporal Video Grounding Capability .....  | 9912 |
| <i>Jiankang Wang, Zhihan Zhang, Zhihang Liu, Yang Li, Jiannan Ge, Hongtao Xie, Yongdong Zhang</i>  |      |

|   |       |
|---|-------|
| SD-PSFNet: Sequential and Dynamic Point Spread Function Network for Image Deraining .....   | 9921  |
| <i>Jiayu Wang, Haoyu Bian, Haoran Sun, Shaoning Zeng</i>  |       |
| SEA-PACE: Semi-Supervised Underwater Image Enhancement Via Gaussian Process-Assisted Self-Paced Learning .....                          | 9930  |
| <i>Jingyang Wang, Hengyue Bi, Jingchao Cao, Feng Gao, Junyu Dong</i>  |       |
| Taming Cascaded Mixture-Of-Experts for Modality-Missing Multi-Modal Salient Object Detection .....                                      | 9939  |
| <i>Kunpeng Wang, Feifan Sun, Keke Chen</i>  |       |
| Temporal and Spatial Representation Learning for Multimodal Low-Beam 3D Object Detection .....  | 9948  |
| <i>Lin Wang, Shiliang Sun, Jing Zhao</i>  |       |
| PointChain: Learning Generalizable Point Cloud Representations Via Structural Chain Modeling .....                                      | 9957  |
| <i>Luyao Wang, Chuxin Wang, Qiao Li, Tianzhu Zhang</i>  |       |
| FantasyTalking2: Timestep-Layer Adaptive Preference Optimization for Audio-Driven Portrait Animation .....                              | 9966  |
| <i>Mengchao Wang, Wang Qiang, Fan Jiang, Mu Xu</i>  |       |
| Modality-Aware Bias Mitigation and Invariance Learning for Unsupervised Visible-Infrared Person Re-Identification .....                 | 9975  |
| <i>Menglin Wang, Xiaojin Gong, Jiachen Li, Genlin Ji</i>  |       |
| PortraitSR: Artist-Inspired Prior Learning for Progressive Face Super-Resolution .....  | 9984  |
| <i>Miaoqing Wang, Jiaxu Leng, Shuang Li, Changjiang Kuang, Long Sun</i>   |       |
| MotionPhysics: Learnable Motion Distillation for Text-Guided Simulation .....   | 9993  |
| <i>Miaowei Wang, Jakub Zadrozny, Oisin Mac Aodha, Amir Vaxman</i>   |       |
| CAST-LUT: Tokenizer-Guided HSV Look-Up Tables for Purple Flare Removal .....  | 10002 |
| <i>Pu Wang, Shuning Sun, Jialang Lu, Chen Wu, Zhihua Zhang, Youshan Zhang, Chenggang Shan, Dianjie Lu, Guijuan Zhang, Zhuoran Zheng</i> |       |
| Adversarial Fair Incomplete Multi-View Clustering .....   | 10011 |
| <i>Qianqian Wang, Haiming Xu, Wei Feng, Quanxue Gao</i>   |       |
| Compression Artifacts Removal for VVC with Frequency Domain Mixture of Experts Network .....  | 10020 |
| <i>Qijun Wang, Kang Wang, Jun Wang</i>  |       |
| BokehCrafter: Taming Video Diffusion Models for Controllable Bokeh Rendering .....  | 10029 |
| <i>Qiwen Wang, Liao Shen, Jiaqi Li, Tianqi Liu, Huiqiang Sun, Zihao Huang, Yachuan Huang, Xianrui Luo, Zhiguo Cao</i>                   |       |
| D3-RSMDE: 40× Faster and High-Fidelity Remote Sensing Monocular Depth Estimation .....  | 10038 |
| <i>Ruizhi Wang, Weihai Li, Zunlei Feng, Haofei Zhang, Mingli Song, Jiayu Wang, Jie Song, Li Sun</i>                                     |       |
| PC-Flow: Preference Alignment in Flow Matching Via Classifier .....   | 10047 |
| <i>Shaomeng Wang, He Wang, Longquan Dai, Jinhui Tang</i>  |       |
| EC-MVSNet: Enhanced Cascaded Multi-View Stereo with Cross-Scale Relevance Integration .....   | 10056 |
| <i>Shaoqian Wang, Jiadai Sun, Bin Fan, Qiang Wang, Bin Lu, Yuchao Dai</i>   |       |
| Exploring Modality-Aware Fusion and Decoupled Temporal Propagation for Multi-Modal Object Tracking .....                                | 10065 |
| <i>Shilei Wang, Pujian Lai, Dong Gao, Jifeng Ning, Gong Cheng</i>   |       |

|   |       |
|---|-------|
| MonoDream: Monocular Vision-Language Navigation with Panoramic Dreaming .....   | 10074 |
| <i>Shuo Wang, Yongcai Wang, Zhaoxin Fan, Yucheng Wang, Maiyue Chen, Kaihui Wang, Zhizhong Su, Wanting Li, Xudong Cai, Yeying Jin, Deying Li</i> |       |
| ObjecTok: Learning Holistic and Robust Object Tokens for MLLMs .....  | 10083 |
| <i>Sihan Wang, Xiyao Liu, Lianqing Liu, Zhi Han</i>   |       |
| PipeDiT: Accelerating Diffusion Transformers in Video Generation with Task Pipelining and Model Decoupling .....                                | 10092 |
| <i>Sijie Wang, Qiang Wang, Shaohuai Shi</i>   |       |
| CaTFormer: Causal Temporal Transformer with Dynamic Contextual Fusion for Driving Intention Prediction .....                                    | 10101 |
| <i>Sirui Wang, Zhou Guan, Bingxi Zhao, Tongjia Gu, Jie Liu</i>  |       |
| GEWDiff: Geometric Enhanced Wavelet-Based Diffusion Model for Hyperspectral Image Super-Resolution.....   | 10109 |
| <i>Sirui Wang, Jiang He, Natàlia Blasco Andreo, Xiao Xiang Zhu</i>  |       |
| Slender3D: Curve-Guided Multi-View Reconstruction of Slender Structures .....   | 10118 |
| <i>Suqin Wang, Zeyi Wang, Min Shi, Zhaoxin Li, Qi Wang, Xiujian Chai, Dengming Zhu</i>  |       |
| You Only Need One Stage: Novel-View Synthesis from a Single Blind Face Image .....  | 10127 |
| <i>Taoyue Wang, Xiang Zhang, Xiaotian Li, Huiyuan Yang, Lijun Yin</i>   |       |
| Towards 3D Object-Centric Feature Learning for Semantic Scene Completion.....   | 10136 |
| <i>Weihua Wang, Yubo Cui, Xiangru Lin, Zhiheng Li, Zheng Fang</i>   |       |
| Debiased Multiplex Tokenizer for Efficient Map-Free Visual Relocalization.....  | 10145 |
| <i>Wenshuai Wang, Hong Liu, Shengquan Li, Peifeng Jiang, Runwei Ding</i>  |       |
| Semore: VLM-Guided Enhanced Semantic Motion Representations for Visual Reinforcement Learning .....   | 10154 |
| <i>Wentao Wang, Chunyang Liu, Kehua Sheng, Bo Zhang, Yan Wang</i>   |       |
| Noisy Correspondence Learning with Modality Gap Direction Correction .....  | 10163 |
| <i>Wuyuqing Wang, Zeyuan Gu, Erkun Yang</i>   |       |
| When Person Re-Identification Meets Event Camera: A Benchmark Dataset and an Attribute-Guided Re-Identification Framework.....                  | 10172 |
| <i>Xiao Wang, Qian Zhu, Shujian Wu, Bo Jiang, Shiliang Zhang</i>  |       |
| Blur-Robust Detection Via Feature Restoration: An End-To-End Framework for Prior-Guided Infrared UAV Target Detection.....                      | 10181 |
| <i>Xiaolin Wang, Houzhang Fang, Qingshan Li, Lu Wang, Yi Chang, Luxin Yan</i>   |       |
| Spectrally Adaptive Channel-Aware Unrolling Network for Compressed Sensing.....   | 10190 |
| <i>Xiaoyang Wang, Hongping Gan</i>  |       |
| UMNet: Uncertainty-Guided Memory Network for Hyperspectral Pansharpening .....  | 10199 |
| <i>Xiaozheng Wang, Yong Yang, Shuying Huang, Nayu Liu, Ziyang Liu</i>   |       |
| Bi-VLM: Binary Post-Training Quantization for Vision-Language Models .....  | 10207 |
| <i>Xijun Wang, Rayyan Abdalla, Junyun Huang, Chengyuan Zhang, Ruiqi Xian, Dinesh Manocha</i>  |       |

|  |       |
|--|-------|
| Generating Attribute-Aware Human Motions from Textual Prompt .....   | 10216 |
| <i>Xinghan Wang, Kun Xu, Fei Li, Cao Sheng, Jiazhong Yu, Yadong Mu</i>   |       |
| PointSLAM++: Robust Dense Neural Gaussian Point Cloud-Based SLAM .....   | 10225 |
| <i>Xu Wang, Boyao Han, Xiaojun Chen, Ying Liu, Ruihui Li</i>   |       |
| X-MoGen: Unified Motion Generation Across Humans and Animals .....   | 10234 |
| <i>Xuan Wang, Kai Ruan, Liyang Qian, Guo Zhi Zhi, Chang Su, Gaoang Wang</i>  |       |
| InterCoser: Interactive 3D Character Creation with Disentangled Fine-Grained Features.....   | 10243 |
| <i>Yi Wang, Jian Ma, Zhuo Su, Guidong Wang, Jingyu Yang, Yu-Kun Lai, Kun Li</i>  |       |
| Deeply Seeking Boundary for Lunar Regolith Segmentation .....  | 10252 |
| <i>Yifeng Wang, Lingxin Wang, Lu Zhang, Yang Li, Chao Xu, Weiwei Zhang, Junyue Tang, Yanhong Zheng, Yong Pang, Shengyuan Jiang, Yi Zhao, Zongquan Deng</i> |       |
| ID-Splat: Propagating Object Identities for Segmenting 3D Aerial-View Scenes.....  | 10261 |
| <i>Yijing Wang, Xu Tang, Xiangrong Zhang, Jingjing Ma</i>  |       |
| Radar-APLANC: Unsupervised Radar-Based Heartbeat Sensing Via Augmented Pseudo-Label and Noise Contrast .....   | 10270 |
| <i>Ying Wang, Zhaodong Sun, Xu Cheng, Zuxian He, Xiaobai Li</i>  |       |
| TileGS: Adaptive Gaussian Densification Through Tile-Guided Perceptual Analysis .....  | 10279 |
| <i>Yiwen Wang, Ran Yi, Lizhuang Ma</i>   |       |
| A Pseudo-Label Optimization Method Based on Polar Coordinate Modeling and Prior Constraints .....  | 10288 |
| <i>Yudi Wang, Hailan Shen, Yixiao Fu, Yuqi Li, Zeshi Lu, Zailiang Chen</i>   |       |
| PATexGS: Perceptual-Adaptive Texture Scheduling for Visual Coherence in Textured Gaussian Splatting .....  | 10297 |
| <i>Yuesong Wang, Dounian Ma, Xiaoyu Chen, Tao Guan</i>   |       |
| ASCD: Attention-Steerable Contrastive Decoding for Reducing Hallucination in MLLM .....  | 10306 |
| <i>Yujun Wang, Jinhe Bi, Soren Pirk, Yunpu Ma</i>  |       |
| Topology-Aware Vision Transformers for Enhanced Scene Recognition.....   | 10315 |
| <i>Yunxi Wang, Shuaiyu Liu, Qiling Li, Yazhou Ren, Xiaorong Pu</i>   |       |
| TIME: Temporal-Sensitive Multi-Dimensional Instruction Tuning and Robust Benchmarking for Video-LLMs .....   | 10323 |
| <i>Yunxiao Wang, Meng Liu, Wenqi Liu, Xuemeng Song, Bin Wen, Fan Yang, Tingting Gao, Di Zhang, Guorui Zhou, Liqiang Nie</i>                                |       |
| Personalize Your Gaussian: Consistent 3D Scene Personalization from a Single Image .....   | 10332 |
| <i>Yuxuan Wang, Xuanyu Yi, Qingshan Xu, Yuan Zhou, Long Chen, Hanwang Zhang</i>  |       |
| MCGS: Markov Chain Gaussian Splatting for Dynamic Scenes Reconstruction .....  | 10341 |
| <i>Yuzhong Wang, Wenmin Wang, Shixiong Zhang, Xinxing Yu, Zhongheng Chen</i>   |       |
| Verb Mirage: Unveiling and Assessing Verb Concept Hallucinations in Multimodal Large Language Models .....   | 10349 |
| <i>Zehao Wang, Xinpeng Liu, Yudonglin Zhang, Xiaoqian Wu, Zhou Fang, Yifan Fang, Junfu Pu, Cewu Lu, Yong-Lu Li</i>   |       |
| Minute-Long Videos with Dual Parallelisms.....   | 10358 |
| <i>Zeqing Wang, Bowen Zheng, Xingyi Yang, Zhenxiong Tan, Yuecong Xu, Xinchao Wang</i>  |       |

|  |       |
|--|-------|
| Difficulty Controlled Diffusion Model for Synthesizing Effective Training Data .....   | 10367 |
| <i>Zerun Wang, Jiafeng Mao, Xueting Wang, Toshihiko Yamasaki</i>   |       |
| Breaking Task Boundaries: A Unified Model for 3D Medical Image Fusion and Segmentation<br>Guided by Manifold Perspective .....   | 10376 |
| <i>Zeyu Wang, Jiayu Wang, Haiyu Song</i>   |       |
| SigFusion: Unified Signal-Level Self-Supervised Learning Paradigm for Image Fusion.....  | 10385 |
| <i>Zeyu Wang, Jiawei Feng, Jiayu Wang, Pengjie Wang, Haiyu Song</i>  |       |
| Fine-Grained Generalization Via Structuralizing Concept and Feature Space into Commonality,<br>Specificity and Confounding ..... | 10394 |
| <i>Zhen Wang, Jiaojiao Zhao, Qilong Wang, Yongfeng Dong, Wenlong Yu</i>  |       |
| SpatioTemporal Difference Network for Video Depth Super-Resolution .....   | 10403 |
| <i>Zhengxue Wang, Yuan Wu, Xiang Li, Zhiqiang Yan, Jian Yang</i>   |       |
| DiffusionPose: Markov-Optimized Diffusion Model for Human Pose Estimation .....  | 10412 |
| <i>Zhigang Wang, Zhenguang Liu, Shaojing Fan, Sifan Wu, Yingying Jiao</i>  |       |
| Beyond Single-Point Perturbation: A Hierarchical, Manifold-Aware Approach to Diffusion Attacks .....                             | 10421 |
| <i>Zhijie Wang, Lin Wang, Zhenyu Wen, Cong Wang</i>  |       |
| T-GVC: Trajectory-Guided Generative Video Coding at Ultra-Low Bitrates .....   | 10430 |
| <i>Zhitao Wang, Hengyu Man, Wenrui Li, Xingtao Wang, Xiaopeng Fan, Debin Zhao</i>  |       |

**Author Index**