



We've just launched a new service: our brand new **dblp SPARQL query service**. Read more about it in our latest blog post or try out some of the SPARQL queries linked on the dblp web pages below.



[+] CVPR 2024: Seattle, WA, USA

> Home > Conferences and Workshops > CVPR

Dagstuhl

- IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024, Seattle, WA, USA, June 16-22, 2024. IEEE 2024, ISBN 979-8-3503-5300-6
- Saurabh Saini, P. J. Narayanan: **Specularity Factorization for Low-Light Enhancement.** 1-12
- Yuyi Liu, Xinhang Song, Weijie Li, Xiaohan Wang, Shuqiang Jiang: **A Category Agnostic Model for Visual Rearrangement.** 1-10
- Yixuan Zhu, Wenliang Zhao, Ao Li, Yansong Tang, Jie Zhou, Jiwen Lu: **FlowIE: Efficient Image Enhancement via Rectified Flow.** 13-22
- Guoqiang Liang, Kanghao Chen, Hangyu Li, Yunfan Lu, Lin Wang: **Towards Robust Event-guided Low-Light Image Enhancement: A Large-Scale Real-World Event-Image Dataset and Novel Approach.** 23-33
- Zhilin Huang, Quanmin Liang, Yijie Yu, Chujun Qin, Xiawu Zheng, Kai Huang, Zikun Zhou, Wenming Yang: **Bilateral Event Mining and Complementary for Event Stream Super-Resolution.** 34-43
- Geunhyuk Youk, Jihyong Oh, Munchurl Kim: **FMA-Net: Flow-Guided Dynamic Filtering and Iterative Feature Refinement with Multi-Attention for Joint Video Super-Resolution and Deblurring.** 44-55
- Yuan Dong, Qi Zuo, Xiaodong Gu, Weihao Yuan, Zhengyi Zhao, Zilong Dong, Liefeng Bo, Qixing Huang: **GPLD3D: Latent Diffusion of 3D Shape Generative Models by Enforcing Geometric and Physical Priors.** 56-66
- Daichi Horita, Naoto Inoue, Kotaro Kikuchi, Kota Yamaguchi, Kiyoharu Aizawa: **Retrieval-Augmented Layout Transformer for Content-Aware Layout Generation.** 67-76
- Dor Verbin, Ben Mildenhall, Peter Hedman, Jonathan T. Barron, Todd E. Zickler, Pratul P. Srinivasan: **Eclipse: Disambiguating Illumination and Materials Using Unintended Shadows.** 77-86
- Bailey Miller, Hanyu Chen, Alice Lai, Ioannis Gkioulekas : **Objects as Volumes: A Stochastic Geometry View of Opaque Solids.** 87-97
- Pakkapon Phongthawee, Worameth Chinchuthakun, Nontaphat Sinsunthithet, Varun Jampani, Amit Raj, Pramook Khungurn, Supasorn Suwanjanakorn: **DiffusionLight: Light Probes for Free by Painting a Chrome Ball.** 98-108
- Zeren Jiang, Chen Guo, Manuel Kaufmann , Tianjian Jiang, Julien Valentin, Otmar Hilliges, Jie Song: **MultiPLY: Reconstruction of Multiple People from Monocular Video in the Wild.** 109-118
- Zhaoxi Chen, Gyeongsik Moon, Kaiwen Guo, Chen Cao, Stanislav Pidhorskyi, Tomas Simon, Rohan Joshi, Yuan Dong, Yichen Xu, Bernardo Pires, He Wen, Lucas Evans, Bo Peng, Julia Buffalini, Autumn Trimble, Kevyn McPhail, Melissa Schoeller, Shoou-I Yu, Javier Romero, Michael Zollhöfer, Yaser Sheikh, Ziwei Liu, Shunsuke Saito: **URHand: Universal Relightable Hands.** 119-129
- Shunsuke Saito, Gabriel Schwartz, Tomas Simon, Junxuan Li, Giljoo Nam: **Relightable Gaussian Codec Avatars.** 130-141
- Xiaoyu Zhan, Jianxin Yang, Yuanqi Li, Jie Guo, Yanwen Guo, Wenping Wang: **Semantic Human Mesh Reconstruction with Textures.** 142-152
- Han Feng, Wenchao Ma, Quankai Gao, Xianwei Zheng, Nan Xue, Huijuan Xu: **Stratified Avatar Generation from Sparse Observations.** 153-163
- Haidong Zhu, Pranav Budhwant, Zhaocheng Zheng, Ram Nevatia: **SEAS: ShapE-Aligned Supervision for Person Re-Identification.** 164-174

[+] SPARQL queries

[–] Refine list

showing all 2,716 records

refine by search term

refine by author

Yu Qiao (24)
Ying Shan (21)
Ziwei Liu (16)
Marc Pollefeys (15)
Ming-Hsuan Yang (15)
Luc Van Gool (15)
Boxin Shi (13)
Matthias Nießner (13)
Dahua Lin (13)
Yi Yang (12)
10,693 more options

refine by access

closed (2,715)



- Qianyu Zhou, Ke-Yue Zhang, Taiping Yao, Xuequan Lu, Shouhong Ding, Lizhuang Ma:
Test-Time Domain Generalization for Face Anti-Spoofing. 175-187
- Binh Minh Le, Simon S. Woo:
Gradient Alignment for Cross-Domain Face Anti-Spoofing. 188-199
- Dingqiang Ye, Chao Fan, Jingzhe Ma, Xiaoming Liu, Shiqi Yu:
BigGait: Learning Gait Representation You Want by Large Vision Models. 200-210
- Xun Lin, Shuai Wang, Rizhao Cai, Yizhong Liu, Ying Fu, Wenzhong Tang, Zitong Yu, Alex C. Kot:
Suppress and Rebalance: Towards Generalized Multi-Modal Face Anti-Spoofing. 211-221
- Ajian Liu, Shuai Xue, Jianwen Gan, Jun Wan, Yanyan Liang, Jiankang Deng, Sergio Escalera, Zhen Lei:
CFPL-FAS: Class Free Prompt Learning for Generalizable Face Anti-Spoofing. 222-232
- Ruijie Quan, Wenguan Wang, Zhibo Tian, Fan Ma, Yi Yang:
Psychometry: An Omnidit Model for Image Reconstruction from Human Brain Activity. 233-243
- Minchul Kim, Yiyang Su, Feng Liu, Anil Jain, Xiaoming Liu:
KeyPoint Relative Position Encoding for Face Recognition. 244-255
- Feng Liu, Minchul Kim, Zhiyuan Ren, Xiaoming Liu:
Distilling CLIP with Dual Guidance for Learning Discriminative Human Body Shape Representation. 256-266
- Leslie Ching Ow Tiong, Dick Sigmund, Chen-Hui Chan, Andrew Beng Jin Teoh:
Flexible Biometrics Recognition: Bridging the Multimodality Gap Through Attention, Alignment and Prompt Tuning. 267-276
- Pei-Kai Huang, Cheng-Hsuan Chiang, Tzu-Hsien Chen, Jun-Xiong Chong, Tyng-Luh Liu, Chiou-Ting Hsu:
One-Class Face Anti-Spoofing via Spoof Cue Map-Guided Feature Learning. 277-286
- Shehreen Azad, Yogesh Singh Rawat:
Activity-Biometrics: Person Identification from Daily Activities. 287-296
- Yuxi Mi, Zhizhou Zhong, Yuge Huang, Jiazen Ji, Jianqing Xu, Jun Wang, Shaoming Wang, Shouhong Ding, Shuigeng Zhou:
Privacy-Preserving Face Recognition Using Trainable Feature Subtraction. 297-307
- Xin Juan, Kaixiong Zhou, Ninghao Liu, Tianlong Chen, Xin Wang:
Molecular Data Programming: Towards Molecule Pseudo-labeling with Systematic Weak Supervision. 308-318
- Ruijie Quan, Wenguan Wang, Fan Ma, Hehe Fan, Yi Yang:
Clustering for Protein Representation Learning. 319-329
- Nathan Mankovich, Gustau Camps-Valls, Tolga Birdal:
Fun with Flags: Robust Principal Directions via Flag Manifolds. 330-340
- Shunsuke Yasuki, Masato Taki:
CAM Back Again: Large Kernel CNNs from a Weakly Supervised Object Localization Perspective. 341-351
- Tsu-Ching Hsiao, Hao-Wei Chen, Hsuan-Kung Yang, Chun-Yi Lee:
Confronting Ambiguity in 6D Object Pose Estimation via Score-Based Diffusion on SE(3). 352-362
- Wooseong Jeong, Kuk-Jin Yoon:
Quantifying Task Priority for Multi-Task Optimization. 363-372
- Chaehyeon Song, Jaeho Shin, Myung-Hwan Jeon, Jongwoo Lim, Ayoung Kim:
Unbiased Estimator for Distorted Conics in Camera Calibration. 373-381
- Xinzhe Wang, Kang Ma, Qiankun Liu, Yunhao Zou, Ying Fu:
Multi-Object Tracking in the Dark. 382-392
- Kaijie Ren, Lei Zhang:
Implicit Discriminative Knowledge Learning for Visible-Infrared Person Re-Identification. 393-402
- Javier Tirado-Garín, Javier Civera:
From Correspondences to Pose: Non-Minimal Certifiably Optimal Relative Pose Without Disambiguation. 403-412
- Hemanth Saratchandran, Sameera Ramasinghe, Simon Lucey:
From Activation to Initialization: Scaling Insights for Optimizing Neural Fields. 413-422
- Ammar Ali, Georgii Gaikov, Denis Rybalchenko, Alexander Chigorin, Ivan Laptev, Sergey Zagoruyko:
PairDETR : Joint Detection and Association of Human Bodies and Faces. 423-432
- Zan Wang, Yixin Chen, Baoxiong Jia, Puhalo Li, Jinlu Zhang, Jingze Zhang, Tengyu Liu, Yixin Zhu, Wei Liang, Siyuan Huang:
Move as you Say, Interact as you can: Language-Guided Human Motion Generation with Scene Affordance. 433-444
- Xinyu Zhan, Lixin Yang, Yifei Zhao, Kangrui Mao, Hanlin Xu, Zenan Lin, Kailin Li, Cewu Lu:
OakInk2 : A Dataset of Bimanual Hands-Object Manipulation in Complex Task Completion. 445-456
- Germán Barquero, Sergio Escalera, Cristina Palmero:
Seamless Human Motion Composition with Blended Positional Encodings. 457-469
- Liao Wang, Kaixin Yao, Chengcheng Guo, Zhirui Zhang, Qiang Hu, Jingyi Yu, Lan Xu, Minye Wu:
VideoRF: Rendering Dynamic Radiance Fields as 2D Feature Video Streams. 470-481
- Han Liang, Jiacheng Bao, Ruichi Zhang, Sihan Ren, Yuecheng Xu, Sibei Yang, Xin Chen, Jingyi Yu, Lan Xu:
OMG: Towards Open-vocabulary Motion Generation via Mixture of Controllers. 482-493



Zicong Fan, Maria Parelli, Maria Eleni Kadoglou, Xu Chen, Muhammed Kocabas, Michael J. Black, Otmar Hilliges: **HOLD: Category-Agnostic 3D Reconstruction of Interacting Hands and Objects from Video.** 494-504

Muhammed Kocabas, Jen-Hao Rick Chang, James Gabriel, Oncel Tuzel, Anurag Ranjan: **HUGS: Human Gaussian Splats.** 505-515

Juze Zhang, Jingyan Zhang, Zining Song, Zhanhe Shi, Chengfeng Zhao, Ye Shi, Jingyi Yu, Lan Xu, Jingya Wang: **HOI-M³: Capture Multiple Humans and Objects Interaction within Contextual Environment.** 516-526

Jihyun Lee, Shunsuke Saito, Giljoo Nam, Minhyuk Sung, Tae-Kyun Kim: **InterHandGen: Two-Hand Interaction Generation via Cascaded Reverse Diffusion.** 527-537

Hsuan-I Ho, Jie Song, Otmar Hilliges: **SiTH: Single-view Textured Human Reconstruction with Image-Conditioned Diffusion.** 538-549

Wenbo Wang, Hsuan-I Ho, Chen Guo, Boxiang Rong, Artur Grigorev, Jie Song, Juan Jose Zarate, Otmar Hilliges: **4D-DRESS: A 4D Dataset of Real-World Human Clothing with Semantic Annotations.** 550-560

Jinglin Xu, Yijie Guo, Yuxin Peng: **FinePOSE: Fine-Grained Prompt-Driven 3D Human Pose Estimation via Diffusion Models.** 561-570

Zhengyi Luo, Jinkun Cao, Rawal Khirodkar, Alexander Winkler, Jing Huang, Kris Kitani, Weipeng Xu: **Real-Time Simulated Avatar from Head-Mounted Sensors.** 571-581

Zhongang Cai, Jianping Jiang, Zhongfei Qing, Xinying Guo, Mingyuan Zhang, Zhengyu Lin, Haiyi Mei, Chen Wei, Ruisi Wang, Wanqi Yin, Liang Pan, Xiangyu Fan, Han Du, Peng Gao, Zhitao Yang, Yang Gao, Jiaqi Li, Tianxiang Ren, Yukun Wei, Xiaogang Wang, Chen Change Loy, Lei Yang, Ziwei Liu: **Digital Life Project: Autonomous 3D Characters with Social Intelligence.** 582-592

Kang Ma, Ying Fu, Chunshui Cao, Saihui Hou, Yongzhen Huang, Dezhi Zheng: **Learning Visual Prompt for Gait Recognition.** 593-603

Wenhao Li, Mengyuan Liu, Hong Liu, Pichao Wang, Jialun Cai, Nicu Sebe: **Hourglass Tokenizer for Efficient Transformer-Based 3D Human Pose Estimation.** 604-613

Dongkai Wang, Shiyu Xuan, Shiliang Zhang: **LocLLM: Exploiting Generalizable Human Keypoint Localization via Large Language Model.** 614-623

Dongkai Wang, Shiliang Zhang: **Spatial-Aware Regression for Keypoint Localization.** 624-633

Liangxiao Hu, Hongwen Zhang, Yuxiang Zhang, Boyao Zhou, Boning Liu, Shengping Zhang, Liqiang Nie: **GaussianAvatar: Towards Realistic Human Avatar Modeling from a Single Video via Animatable 3D Gaussians.** 634-644

Mengcheng Li, Hongwen Zhang, Yuxiang Zhang, Ruizhi Shao, Tao Yu, Yebin Liu: **HHMR: Holistic Hand Mesh Recovery by Enhancing the Multimodal Controllability of Graph Diffusion Models.** 645-654

Qi Fang, Yinghui Fan, Yanjun Li, Junting Dong, Dingwei Wu, Weidong Zhang, Kang Chen: **Capturing Closely Interacted Two-Person Motions with Reaction Priors.** 655-665

Ziqiao Peng, Wentao Hu, Yue Shi, Xiangyu Zhu, Xiaomei Zhang, Hao Zhao, Jun He, Hongyan Liu, Zhaoxin Fan: **SyncTalk: The Devil is in the Synchronization for Talking Head Synthesis.** 666-676

Ruicong Liu, Takehiko Ohkawa, Mingfang Zhang, Yoichi Sato: **Single-to-Dual-View Adaptation for Egocentric 3D Hand Pose Estimation.** 677-686

Canyu Zhang, Youbao Tang, Ning Zhang, Ruei-Sung Lin, Mei Han, Jing Xiao, Song Wang: **Bidirectional Autoregressive Diffusion Model for Dance Generation.** 687-696

Yuxuan Han, Junfeng Lyu, Feng Xu: **High-Quality Facial Geometry and Appearance Capture at Home.** 697-707

Ziwei Liao, Jialiang Zhu, Chunyu Wang, Han Hu, Steven L. Waslander: **Multiple View Geometry Transformers for 3D Human Pose Estimation.** 708-717

Jingbo Wang, Zhengyi Luo, Ye Yuan, Yixuan Li, Bo Dai: **PACER+: On-Demand Pedestrian Animation Controller in Driving Scenarios.** 718-728

Chengfeng Zhao, Juze Zhang, Jiashen Du, Ziwei Shan, Junye Wang, Jingyi Yu, Jingya Wang, Lan Xu: **I'M HOI: Inertia-Aware Monocular Capture of 3D Human-Object Interactions.** 729-741

Xihe Yang, Xingyu Chen, Daiheng Gao, Shaohui Wang, Xiaoguang Han, Baoyuan Wang: **HAVE-FUN: Human Avatar Reconstruction from Few-Shot Unconstrained Images.** 742-752

Inhwan Bae, Junoh Lee, Hae-Gon Jeon: **Can Language Beat Numerical Regression? Language-Based Multimodal Trajectory Prediction.** 753-766

Hiroyasu Akada, Jian Wang, Vladislav Golyanik, Christian Theobalt: **3D Human Pose Perception from Egocentric Stereo Videos.** 767-776

Jian Wang, Zhe Cao, Diogo C. Luvizon, Lingjie Liu, Kripasindhu Sarkar, Danhang Tang, Thabo Beeler, Christian Theobalt: **Egocentric Whole-Body Motion Capture with FisheyeViT and Diffusion-Based Motion Refinement.** 777-787

Arthur Moreau, Jifei Song, Helisa Dhamo, Richard Shaw, Yiren Zhou, Eduardo Pérez-Pellitero: **Human Gaussian Splatting: Real-Time Rendering of Animatable Avatars.** 788-798



- Xiaozheng Zheng, Chao Wen, Zhuo Su, Zeran Xu, Zhaoju Li, Yang Zhao, Zhou Xue:
OHTA: One-shot Hand Avatar via Data-driven Implicit Priors. 799-810
- Wenfeng Song, Xinyu Zhang, Shuai Li, Yang Gao, Aimin Hao, Xia Hau, Chenglizhao Chen, Ning Li, Hong Qin:
HOIAnimator: Generating Text-Prompt Human-Object Animations Using Novel Perceptive Diffusion Models. 811-820
- Wenfeng Song, Xingliang Jin, Shuai Li, Chenglizhao Chen, Aimin Hao, Xia Hou, Ning Li, Hong Qin:
Arbitrary Motion Style Transfer with Multi-Condition Motion Latent Diffusion Model. 821-830
- Yan-Kang Wang, Chengyi Xing, Yi-Lin Wei, Xiao-Ming Wu, Wei-Shi Zheng:
Single-View Scene Point Cloud Human Grasp Generation. 831-841
- Taeho Kang, Youngki Lee:
Attention-Propagation Network for Egocentric Heatmap to 3D Pose Lifting. 842-851
- Jieming Cui, Tengyu Liu, Nian Liu, Yaodong Yang, Yixin Zhu, Siyuan Huang:
AnySkill: Learning Open-Vocabulary Physical Skill for Interactive Agents. 852-862
- Zekun Qian, Ruize Han, Wei Feng, Song Wang:
From a Bird's Eye View to See: Joint Camera and Subject Registration without the Camera Calibration. 863-873
- Peng Dai, Yang Zhang, Tao Liu, Zhen Fan, Tianyuan Du, Zhuo Su, Xiaozheng Zheng, Zeming Li:
HMD-Poser: On-Device Real-time Human Motion Tracking from Scalable Sparse Observations. 874-884
- Xingyu Ren, Jiankang Deng, Yuhao Cheng, Jia Guo, Chao Ma, Yichao Yan, Wenhan Zhu, Xiaokang Yang:
Monocular Identity-Conditioned Facial Reflectance Reconstruction. 885-895
- Ye Yuan, Xuetong Li, Yangyi Huang, Shalini De Mello, Koki Nagano, Jan Kautz, Umar Iqbal:
GAvatar: Animatable 3D Gaussian Avatars with Implicit Mesh Learning. 896-905
- Anastasis Stathopoulos, Ligong Han, Dimitris N. Metaxas:
Score-Guided Diffusion for 3D Human Recovery. 906-915
- Yuhao Cheng, Zhuo Chen, Xingyu Ren, Wenhan Zhu, Zhengqin Xu, Di Xu, Changpeng Yang, Yichao Yan:
3D-Aware Face Editing via Warping-Guided Latent Direction Learning. 916-926
- Markos Diomataris, Nikos Athanasiou, Omid Taheri, Xi Wang, Otmar Hilliges, Michael J. Black:
WANDR: Intention-guided Human Motion Generation. 927-936
- Qing Yu, Mikihiro Tanaka, Kent Fujiwara:
Exploring Vision Transformers for 3D Human Motion-Language Models with Motion Patches. 937-946
- Nilesh Kulkarni, Davis Rempe, Kyle Genova, Abhijit Kundu, Justin Johnson, David Fouhey, Leonidas J. Guibas:
NIFTY: Neural Object Interaction Fields for Guided Human Motion Synthesis. 947-957
- Yukang Cao, Yan-Pei Cao, Kai Han, Ying Shan, Kwan-Yee K. Wong:
DreamAvatar: Text-and-Shape Guided 3D Human Avatar Generation via Diffusion Models. 958-968
- Kangwei Yan, Fei Wang, Bo Qian, Han Ding, Jinsong Han, Xing Wei:
Person-in-WiFi 3D: End-to-End Multi-Person 3D Pose Estimation with Wi-Fi. 969-978
- Yuan Xu, Xiaoxuan Ma, Jiajun Su, Wentao Zhu, Yu Qiao, Yizhou Wang:
ScoreHypo: Probabilistic Human Mesh Estimation with Hypothesis Scoring. 979-989
- Zhen Xu, Sida Peng, Chen Geng, Linzhan Mou, Zihan Yan, Jiaming Sun, Hujun Bao, Xiaowei Zhou:
Relightable and Animatable Neural Avatar from Sparse-View Video. 990-1000
- Evonne Ng, Javier Romero, Timur M. Bagautdinov, Shaojie Bai, Trevor Darrell, Angjoo Kanazawa, Alexander Richard:
From Audio to Photoreal Embodiment: Synthesizing Humans in Conversations. 1001-1010
- Buzhen Huang, Chen Li, Chongyang Xu, Liang Pan, Yangang Wang, Gim Hee Lee:
Closely Interactive Human Reconstruction with Proxemics and Physics-Guided Adaption. 1011-1021
- Jijie He, Wenwu Yang:
Video-Based Human Pose Regression via Decoupled Space-Time Aggregation. 1022-1031
- Chengyang Hu, Ke-Yue Zhang, Taiping Yao, Shouhong Ding, Lizhuang Ma:
Rethinking Generalizable Face Anti-Spoofing via Hierarchical Prototype-Guided Distribution Refinement in Hyperbolic Space. 1032-1041
- Xiaoning Sun, Huaijiang Sun, Bin Li, Dong Wei, Weiqing Li, Jianfeng Lu:
MoML: Online Meta Adaptation for 3D Human Motion Prediction. 1042-1051
- Fengyuan Yang, Kerui Gu, Angela Yao:
KITRO: Refining Human Mesh by 2D Clues and Kinematic-tree Rotation. 1052-1061
- Inhee Lee, Byungjun Kim, Hanbyul Joo:
Guess The Unseen: Dynamic 3D Scene Reconstruction from Partial 2D Glimpses. 1062-1071
- Hyunsoo Cha, Byungjun Kim, Hanbyul Joo:
PEGASUS: Personalized Generative 3D Avatars with Composable Attributes. 1072-1081
- Sichen Chen, Yingyi Zhang, Siming Huang, Ran Yi, Ke Fan, Ruixin Zhang, Peixian Chen, Jun Wang, Shouhong Ding, Lizhuang Ma:
SDPose: Tokenized Pose Estimation via Circulation-Guide Self-Distillation. 1082-1090



- Jiye Lee, Hanbyul Joo:
Mocap Everyone Everywhere: Lightweight Motion Capture with Smartwatches and a Head-Mounted Camera. 1091-1100
- Yixuan Zhu, Ao Li, Yansong Tang, Wenliang Zhao, Jie Zhou, Jiwen Lu:
DPMesh: Exploiting Diffusion Prior for Occluded Human Mesh Recovery. 1101-1110
- Jiapeng Tang, Angela Dai, Yinyu Nie, Lev Markhasin, Justus Thies, Matthias Nießner:
DPHMs: Diffusion Parametric Head Models for Depth-Based Tracking. 1111-1122
- Jihua Peng, Yanghong Zhou, P. Y. Mok:
KTPFormer: Kinematics and Trajectory Prior Knowledge-Enhanced Transformer for 3D Human Pose Estimation. 1123-1132
- Jongwook Choi, Taehoon Kim, Yonghyun Jeong, Seungryul Baek, Jongwon Choi:
Exploiting Style Latent Flows for Generalizing Deepfake Video Detection. 1133-1143
- Haiyang Liu, Zihao Zhu, Giorgio Becherini, Yichen Peng, Mingyang Su, You Zhou, Xuefei Zhe, Naoya Iwamoto, Bo Zheng, Michael J. Black:
EMAGE: Towards Unified Holistic Co-Speech Gesture Generation via Expressive Masked Audio Gesture Modeling. 1144-1154
- Yiteng Xu, Kecheng Ye, Xiao Han, Yiming Ren, Xinge Zhu, Yuexin Ma:
A Unified Framework for Human-centric Point Cloud Video Understanding. 1155-1164
- Haokai Pang, Heming Zhu, Adam Kortylewski, Christian Theobalt, Marc Habermann:
ASH: Animatable Gaussian Splats for Efficient and Photoreal Human Rendering. 1165-1175
- Andrey Davydov, Martin Engelberge, Mathieu Salzmann, Pascal Fua:
CLOAF: CoLlisiOn-Aware Human Flow. 1176-1185
- Christen Millerdurair, Hiroyasu Akada, Jian Wang, Diogo C. Luvizon, Christian Theobalt, Vladislav Golyanik:
EventEgo3D: 3D Human Motion Capture from Egocentric Event Streams. 1186-1195
- Jakub Paplhám, Vojtech Franc:
A Call to Reflect on Evaluation Practices for Age Estimation: Comparative Analysis of the State-of-the-Art and a Unified Benchmark. 1196-1205
- Ashwath Shetty, Marc Habermann, Guoxing Sun, Diogo C. Luvizon, Vladislav Golyanik, Christian Theobalt:
Holoported Characters: Real-Time Free-Viewpoint Rendering of Humans from Sparse RGB Cameras. 1206-1215
- Yizhou Zhao, Tuanfeng Yang Wang, Bhiksha Raj, Min Xu, Jimei Yang, Chun-Hao Paul Huang:
Synergistic Global-Space Camera and Human Reconstruction from Videos. 1216-1226
- Felix Taubner, Prashant Raina, Mathieu Tuli, Eu Wern Teh, Chul Lee, Jinmiao Huang:
3D Face Tracking from 2D Video through Iterative Dense UV to Image Flow. 1227-1237
- Mingyu Zhou, Rakib Hyder, Ziwei Xuan, Guojun Qi:
UltrAvatar: A Realistic Animatable 3D Avatar Diffusion Model with Authenticity Guided Textures. 1238-1248
- Zhangsihao Yang, Mingyu Zhou, Mengyi Shan, Bingbing Wen, Ziwei Xuan, Mitch Hill, Junjie Bai, Guo-Jun Qi, Yalin Wang:
OmniMotionGPT: Animal Motion Generation with Limited Data. 1249-1259
- Yunjie Wu, Yapeng Meng, Zhipeng Hu, Lincheng Li, Haoqian Wu, Kun Zhou, Weiwei Xu, Xin Yu:
Text-Guided 3D Face Synthesis - From Generation to Editing. 1260-1269
- Zihan Wang, Siyang Song, Cheng Luo, Songhe Deng, Weicheng Xie, Linlin Shen:
Multi-Scale Dynamic and Hierarchical Relationship Modeling for Facial Action Units Recognition. 1270-1280
- Yiming Ren, Xiao Han, Chengfeng Zhao, Jingya Wang, Lan Xu, Jingyi Yu, Yuexin Ma:
LiveHPS: LiDAR-Based Scene-Level Human Pose and Shape Estimation in Free Environment. 1281-1291
- Chao Xu, Yang Liu, Jiazheng Xing, Weida Wang, Mingze Sun, Jun Dan, Tianxin Huang, Siyuan Li, Zhi-Qi Cheng, Ying Tai, Baigui Sun:
FaceChain-ImagineID: Freely Crafting High-Fidelity Diverse Talking Faces from Disentangled Audio. 1292-1302
- Yuchen Pan, Junjun Jiang, Kui Jiang, Zhihao Wu, Keyuan Yu, Xianming Liu:
OpticalIDR: A Deep Optical Imaging Model for Privacy-Protective Depression Recognition. 1303-1312
- Kejia Yin, Varshanth S. Rao, Ruowei Jiang, Xudong Liu, Parham Aarabi, David B. Lindell:
SCE-MAE: Selective Correspondence Enhancement with Masked Autoencoder for Self-Supervised Landmark Estimation. 1313-1322
- Sai Kumar Dwivedi, Yu Sun, Priyanka Patel, Yao Feng, Michael J. Black:
TokenHMR: Advancing Human Mesh Recovery with a Tokenized Pose Representation. 1323-1333
- Korrawe Karunratanakul, Konpat Preechakul, Emre Aksan, Thabo Beeler, Supasorn Suwajanakorn, Siyu Tang:
Optimizing Diffusion Noise Can Serve As Universal Motion Priors. 1334-1345
- Luyang Zhu, Yingwei Li, Nan Liu, Hao Peng, Dawei Yang, Ira Kemelmacher-Shlizerman:
M&M VTO: Multi-Garment Virtual Try-On and Editing. 1346-1356
- Zixiang Zhou, Yu Wan, Baoyuan Wang:
AvatarGPT: All-in-One Framework for Motion Understanding, Planning, Generation and Beyond. 1357-1366
- Zhishan Zhou, Shihao Zhou, Zhi Lv, Minqiang Zou, Yao Tang, Jiajun Liang:
A Simple Baseline for Efficient Hand Mesh Reconstruction. 1367-1376



- Zhouyingcheng Liao, Vladislav Golyanik, Marc Habermann, Christian Theobalt:
VINECS: Video-based Neural Character Skinning. 1377-1387
- Muhammad Hamza Mughal, Rishabh Dabral, Ikhsanul Habibie, Lucia Donatelli, Marc Habermann, Christian Theobalt:
ConvoFusion: Multi-Modal Conversational Diffusion for Co-Speech Gesture Synthesis. 1388-1398
- Hanchao Liu, Xiaohang Zhan, Shaoli Huang, Tai-Jiang Mu, Ying Shan:
Programmable Motion Generation for Open-Set Motion Control Tasks. 1399-1408
- Yiwei Bao, Feng Lu:
From Feature to Gaze: A Generalizable Replacement of Linear Layer for Gaze Estimation. 1409-1418
- Yiwei Bao, Feng Lu:
Unsupervised Gaze Representation Learning from Multi-view Face Images. 1419-1428
- Muxin Zhang, Qiao Feng, Zhuo Su, Chao Wen, Zhou Xue, Kun Li:
Joint2Human: High-quality 3D Human Generation via Compact Spherical Embedding of 3D Joints. 1429-1438
- Akash Sengupta, Thimo Alldieck, Nikos Kolotouros, Enric Corona, Andrei Zanfir, Cristian Sminchisescu:
DiffHuman: Probabilistic Photorealistic 3D Reconstruction of Humans. 1439-1449
- Youliang Zhang, Wenxuan Liu, Danni Xu, Zhuo Zhou, Zheng Wang:
Bi-Causal: Group Activity Recognition via Bidirectional Causality. 1450-1459
- Caoyuan Ma, Yu-Lun Liu, Zhixiang Wang, Wu Liu, Xinchen Liu, Zheng Wang:
HumanNeRF-SE: A Simple yet Effective Approach to Animate HumanNeRF with Diverse Poses. 1460-1470
- Haoyang Ge, Qiao Feng, Hailong Jia, Xiongzheng Li, Xiangjun Yin, You Zhou, Jingyu Yang, Kun Li:
LPSNet: End-to-End Human Pose and Shape Estimation with Lensless Imaging. 1471-1480
- Zhongcong Xu, Jianfeng Zhang, Jun Hao Liew, Hanshu Yan, Jia-Wei Liu, Chenxu Zhang, Jiashi Feng, Mike Zheng Shou:
MagicAnimate: Temporally Consistent Human Image Animation using Diffusion Model. 1481-1490
- Peng Lu, Tao Jiang, Yining Li, Xiangtai Li, Kai Chen, Wenming Yang:
RTMO: Towards High-Performance One-Stage Real-Time Multi-Person Pose Estimation. 1491-1500
- Jiangbei Yue, Baiyi Li, Julien Pettré, Armin Seyfried, He Wang:
Human Motion Prediction Under Unexpected Perturbation. 1501-1511
- Matthieu Armando, Salma Galaaoui, Fabien Baradel, Thomas Lucas, Vincent Leroy, Romain Brégier, Philippe Weinzaepfel, Grégory Rogez:
Cross-View and Cross-Pose Completion for 3D Human Understanding. 1512-1523
- Ronghui Li, Yuxiang Zhang, Yachao Zhang, Hongwen Zhang, Jie Guo, Yan Zhang, Yebin Liu, Xiu Li:
Lodge: A Coarse to Fine Diffusion Network for Long Dance Generation Guided by the Characteristic Dance Primitives. 1524-1534
- Taeksoo Kim, Byungjun Kim, Shunsuke Saito, Hanbyul Joo:
GALA: Generating Animatable Layered Assets from a Single Scan. 1535-1545
- Ekkasit Pinyoanuntapong, Pu Wang, Minwoo Lee, Chen Chen:
MMM: Generative Masked Motion Model. 1546-1555
- Yihua Cheng, Yaning Zhu, Zongji Wang, Hongquan Hao, Yongwei Liu, Shiqing Cheng, Xi Wang, Hyung Jin Chang:
What Do You See in Vehicle? Comprehensive Vision Solution for In-Vehicle Gaze Estimation. 1556-1565
- Yifei Liu, Qiong Cao, Yandong Wen, Huaiyang Jiang, Changxing Ding:
Towards Variable and Coordinated Holistic Co-Speech Motion Generation. 1566-1576
- Junuk Cha, Jihyeon Kim, Jae Shin Yoon, Seungryul Baek:
Text2HOI: Text-Guided 3D Motion Generation for Hand-Object Interaction. 1577-1585
- Ren Li, Corentin Dumery, Benoît Guillard, Pascal Fua:
Garment Recovery with Shape and Deformation Priors. 1586-1595
- Kangning Yin, Shihao Zou, Yuxuan Ge, Zheng Tian:
Tri-Modal Motion Retrieval by Learning a Joint Embedding Space. 1596-1605
- Zhijing Shao, Zhaolong Wang, Zhuang Li, Duotun Wang, Xiangru Lin, Yu Zhang, Mingming Fan, Zeyu Wang:
SplattingAvatar: Realistic Real-Time Human Avatars With Mesh-Embedded Gaussian Splatting. 1606-1616
- Jiaqi Liao, Chuanchen Luo, Yinuo Du, Yuxi Wang, Xucheng Yin, Man Zhang, Zhaoxiang Zhang, Junran Peng:
HardMo: A Large-Scale Hardcase Dataset for Motion Capture. 1629-1638
- Zhonglin Sun, Chen Feng, Ioannis Patras, Georgios Tzimiropoulos:
LAFS: Landmark-Based Facial Self-Supervised Learning for Face Recognition. 1639-1649
- Hee Jae Kim, Eshed Ohn-Bar:
Motion Diversification Networks. 1650-1660
- Yannan He, Garvita Tiwari, Tolga Birdal, Jan Eric Lenssen, Gerard Pons-Moll:
NRDF: Neural Riemannian Distance Fields for Learning Articulated Pose Priors. 1661-1671
- Zidu Wang, Xiangyu Zhu, Tianshuo Zhang, Baiqin Wang, Zhen Lei:
3D Face Reconstruction with the Geometric Guidance of Facial Part Segmentation. 1672-1682
- Zhibo Yang, Sounak Mondal, Seoyoung Ahn, Ruoyu Xue, Gregory J. Zelinsky, Minh Hoai, Dimitris Samaras:
Unifying Top-Down and Bottom-Up Scanpath Prediction Using Transformers. 1683-1693



- Fu-Zhao Ou, Chongyi Li, Shiqi Wang, Sam Kwong:
CLIB-FIQA: Face Image Quality Assessment with Confidence Calibration. 1694-1704
- Boeun Kim, Jungho Kim, Hyung Jin Chang, Jin Young Choi:
MoST: Motion Style Transformer Between Diverse Action Contents. 1705-1714
- Yuxiao Liu, Zhe Li, Yebin Liu, Haoqian Wang:
TexVocab: Texture Vocabulary-Conditioned Human Avatars. 1715-1725
- Haitao Yan, Qiongjie Cui, Jiexin Xie, Shijie Guo:
Forecasting of 3D Whole-Body Human Poses with Grasping Objects. 1726-1736
- Nan Jiang, Zhiyuan Zhang, Hongjie Li, Xiaoxuan Ma, Zan Wang, Yixin Chen, Tengyu Liu, Yixin Zhu, Siyuan Huang:
Scaling Up Dynamic Human-Scene Interaction Modeling. 1737-1747
- Jiali Zheng, Rolando Alexandros Potamias, Stefanos Zafeiriou:
Design2Cloth: 3D Cloth Generation from 2D Masks. 1748-1758
- Liang Xu, Yizhou Zhou, Yichao Yan, Xin Jin, Wenhan Zhu, Fengyun Rao, Xiaokang Yang, Wenjun Zeng:
ReGenNet: Towards Human Action-Reaction Synthesis. 1759-1769
- Abdallah Dib, Luiz Gustavo Hafemann, Emeline Got, Trevor Anderson, Amin Fadaeinejad, Rafael M. O. Cruz, Marc-André Carboneau:
MoSAR: Monocular Semi-Supervised Model for Avatar Reconstruction using Differentiable Shading. 1770-1780
- David Ferman, Pablo Garrido, Gaurav Bharaj:
FaceLift: Semi-Supervised 3D Facial Landmark Localization. 1781-1791
- Shengxiang Hu, Huaijiang Sun, Bin Li, Dong Wei, Weiqing Li, Jianfeng Lu:
Fast Adaptation for Human Pose Estimation via Meta-Optimization. 1792-1801
- Jun Xiang, Xuan Gao, Yudong Guo, Juyong Zhang:
FlashAvatar: High-Fidelity Head Avatar with Efficient Gaussian Embedding. 1802-1812
- Tianyu Li, Calvin Qiao, Guanqiao Ren, KangKang Yin, Sehoon Ha:
AAMDM: Accelerated Auto-Regressive Motion Diffusion Model. 1813-1823
- Tao Wang, Lei Jin, Zheng Wang, Jianshu Li, Liang Li, Fang Zhao, Yu Cheng, Li Yuan, Li Zhou, Junliang Xing, Jian Zhao:
SynSP: Synergy of Smoothness and Precision in Pose Sequences Refinement. 1824-1833
- Qingping Sun, Yanjun Wang, Ailing Zeng, Wanqi Yin, Chen Wei, Wenjia Wang, Haiyi Mei, Chi-Sing Leung, Ziwei Liu, Lei Yang, Zhongang Cai:
AiOS: All-in-One-Stage Expressive Human Pose and Shape Estimation. 1834-1843
- Jingbo Zhang , Xiaoyu Li, Qi Zhang, Yanpei Cao, Ying Shan, Jing Liao :
HumanRef: Single Image to 3D Human Generation via Reference-Guided Diffusion. 1844-1854
- Zhi Cen, Huaijin Pi, Sida Peng, Zehong Shen, Minghui Yang, Shuai Zhu, Hujun Bao, Xiaowei Zhou:
Generating Human Motion in 3D Scenes from Text Descriptions. 1855-1866
- Michail Tarasiou, Rolando Alexandros Potamias, Eimear O' Sullivan, Stylianos Ploumpis, Stefanos Zafeiriou:
Locally Adaptive Neural 3D Morphable Models. 1867-1876
- Shaofei Wang, Bozidar Antic, Andreas Geiger, Siyu Tang:
IntrinsicAvatar: Physically Based Inverse Rendering of Dynamic Humans from Monocular Videos via Explicit Ray Tracing. 1877-1888
- Yu Zhang, Songpengcheng Xia, Lei Chu, Jiarui Yang, Qi Wu, Ling Pei:
Dynamic Inertial Poser (DynalIP): Part-Based Motion Dynamics Learning for Enhanced Human Pose Estimation with Sparse Inertial Sensors. 1889-1899
- Chuan Guo, Yuxuan Mu, Muhammad Gohar Javed, Sen Wang, Li Cheng:
MoMask: Generative Masked Modeling of 3D Human Motions. 1900-1910
- Yufei Ye, Abhinav Gupta, Kris Kitani, Shubham Tulsiani:
G-HOP: Generative Hand-Object Prior for Interaction Reconstruction and Grasp Synthesis. 1911-1920
- Pengfei Ren, Yuanyuan Gao, Haifeng Sun, Qi Qi, Jingyu Wang, Jianxin Liao:
Dynamic Support Information Mining for Category-Agnostic Pose Estimation. 1921-1930
- Yuelang Xu, Bengwang Chen, Zhe Li, Hongwen Zhang, Lizhen Wang, Zerong Zheng, Yebin Liu:
Gaussian Head Avatar: Ultra High-Fidelity Head Avatar via Dynamic Gaussians. 1931-1941
- Kiran Chhatre, Radek Danecek, Nikos Athanasiou, Giorgio Becherini, Christopher Peters, Michael J. Black, Timo Bolkart:
Emotional Speech-Driven 3D Body Animation via Disentangled Latent Diffusion. 1942-1953
- Yuxiang Zhang, Hongwen Zhang, Liangxiao Hu, Jiajun Zhang, Hongwei Yi, Shengping Zhang, Yebin Liu:
ProxyCap: Real-Time Monocular Full-Body Capture in World Space via Human-Centric Proxy-to-Motion Learning. 1954-1964
- Roy Kapon, Guy Tevet, Daniel Cohen-Or, Amit H. Bermano:
MAS: Multi-view Ancestral Sampling for 3D Motion Generation Using 2D Diffusion. 1965-1974
- Ziqian Bai, Feitong Tan, Sean Fanello, Rohit Pandey, Mingsong Dou, Shichen Liu, Ping Tan, Yinda Zhang:
Efficient 3D Implicit Head Avatar With Mesh-Anchored Hash Table Blendshapes. 1975-1984
- Vasileios Baltatzis, Rolando Alexandros Potamias, Evangelos Ververas, Guanxiong Sun, Jiankang Deng, Stefanos Zafeiriou:
Neural Sign Actors: A diffusion model for 3D sign language production from text. 1985-1995



- Xiang Deng, Zerong Zheng, Yuxiang Zhang, Jingxiang Sun, Chao Xu, Xiaodong Yang, Lizhen Wang, Yebin Liu:
RAM-Avatar: Real-time Photo-Realistic Avatar from Monocular Videos with Full-body Control. 1996-2007
- Samy Tafasca, Anshul Gupta, Jean-Marc Odobez:
Sharingan: A Transformer Architecture for Multi-Person Gaze Following. 2008-2017
- Yan Zhang, Sergey Prokudin, Marko Mihajlovic, Qianli Ma, Siyu Tang:
Degrees of Freedom Matter: Inferring Dynamics from Point Trajectories. 2018-2028
- Gyeongsik Moon, Weipeng Xu, Rohan Joshi, Chenglei Wu, Takaaki Shiratori:
Authentic Hand Avatar from a Phone Scan via Universal Hand Model. 2029-2038
- Nannan Li, Qing Liu, Krishna Kumar Singh, Yilin Wang, Jianming Zhang, Bryan A. Plummer, Zhe Lin:
UniHuman: A Unified Model For Editing Human Images in the Wild. 2039-2048
- Yuxuan Zhou, Xudong Yan, Zhi-Qi Cheng, Yan Yan, Qi Dai, Xian-Sheng Hua:
BlockGCN: Redefine Topology Awareness for Skeleton-Based Action Recognition. 2049-2058
- Jing Wen, Xiaoming Zhao, Zhongzheng Ren, Alexander G. Schwing, Shenlong Wang:
GoMAvatar: Efficient Animatable Human Modeling from Monocular Video Using Gaussians-on-Mesh. 2059-2069
- Soyong Shin, Juyong Kim, Eni Halilaj, Michael J. Black:
WHAM: Reconstructing World-Grounded Humans with Accurate 3D Motion. 2070-2080
- Zheng Gao, Ioannis Patras:
Self-Supervised Facial Representation Learning with Facial Region Awareness. 2081-2092
- Yao Feng, Jing Lin, Sai Kumar Dwivedi, Yu Sun, Priyanka Patel, Michael J. Black:
ChatPose: Chatting about 3D Human Pose. 2093-2103
- Shiwei Jin, Zhen Wang, Lei Wang, Peng Liu, Ning Bi, Truong Nguyen:
AUEditNet: Dual-Branch Facial Action Unit Intensity Manipulation with Implicit Disentanglement. 2104-2113
- Renshuai Liu, Bowen Ma, Wei Zhang, Zhipeng Hu, Changjie Fan, Tangjie Lv, Yu Ding, Xuan Cheng:
Towards a Simultaneous and Granular Identity-Expression Control in Personalized Face Generation. 2114-2123
- Yanlu Cai, Weizhong Zhang, Yuan Wu, Cheng Jin:
PoseIRM: Enhance 3D Human Pose Estimation on Unseen Camera Settings via Invariant Risk Minimization. 2124-2133
- Haipeng Chen, Kedi Lyu, Zhenguang Liu, Yifang Yin, Xun Yang, Yingda Lyu:
Rethinking Human Motion Prediction with Symplectic Integral. 2134-2143
- Zhenyu Lou, Qiongjie Cui, Haofan Wang, Xu Tang, Hong Zhou:
Multimodal Sense-Informed Forecasting of 3D Human Motions. 2144-2154
- Haodong Zhang, Zhike Chen, Haocheng Xu, Lei Hao, Xiaofei Wu, Songcen Xu, Zhensong Zhang, Yue Wang, Rong Xiong:
Semantics-Aware Motion Retargeting with Vision-Language Models. 2155-2164
- Xingchao Yang, Takafumi Takeomi, Yuki Endo, Yoshihiro Kanamori:
Makeup Prior Models for 3D Facial Makeup Estimation and Applications. 2165-2175
- Yinglong Li, Hongyu Wu, Xiaogang Wang, Qingzhao Qin, Yijiao Zhao, Yong Wang, Aimin Hao:
FaceCom: Towards High-fidelity 3D Facial Shape Completion via Optimization and Inpainting Guidance. 2177-2186
- Xiaoming Li, Xinyu Hou, Chen Change Loy:
When StyleGAN Meets Stable Diffusion: a \mathcal{W} -Adapter for Personalized Image Generation. 2187-2196
- Chandradeep Pokhriya, Ishaan Nikhil Shah, Angela Xing, Zekun Li, Kefan Chen, Avinash Sharma, Srinath Sridhar:
MANUS: Markerless Grasp Capture Using Articulated 3D Gaussians. 2197-2208
- Chengxu Zuo, Yiming Wang, Lishuang Zhan, Shihui Guo, Xinyu Yi, Feng Xu, Yipeng Qin:
Loose Inertial Poser: Motion Capture with IMU-attached Loose-Wear Jacket. 2209-2219
- Prashanth Chandran, Gaspard Zoss:
Anatomically Constrained Implicit Face Models. 2220-2229
- Dayi Tan, Hansheng Chen, Wei Tian, Lu Xiong:
DiffusionRegPose: Enhancing Multi-Person Pose Estimation Using a Diffusion-Based End-to-End Regression Approach. 2230-2239
- Qucheng Peng, Ce Zheng, Chen Chen:
A Dual-Augmentor Framework for Domain Generalization in 3D Human Pose Estimation. 2240-2249
- Ming Yan, Yan Zhang, Shuqiang Cai, Shuqi Fan, Xincheng Lin, Yudi Dai, Siqi Shen, Chenglu Wen, Lan Xu, Yuexin Ma, Cheng Wang:
RELI11D: A Comprehensive Multimodal Human Motion Dataset and Method. 2250-2262
- Xu He, Qiaochu Huang, Zhensong Zhang, Zhiwei Lin, Zhiyong Wu, Sicheng Yang, Minglei Li, Zhiyi Chen, Songcen Xu, Xiaofei Wu:
Co-Speech Gesture Video Generation via Motion-Decoupled Diffusion Model. 2263-2273
- Wencan Cheng, Hao Tang, Luc Van Gool, Jong Hwan Ko:
HandDiff: 3D Hand Pose Estimation with Diffusion on Image-Point Cloud. 2274-2284



- Olaf Dünkel, Tim Salzmann, Florian Pfaff:
Normalizing Flows on the Product Space of SO(3) Manifolds for Probabilistic Human Pose Modeling. 2285-2294
- Haoyu Chen, Hao Tang, Ehsan Adeli, Guoying Zhao:
Towards Robust 3D Pose Transfer with Adversarial Learning. 2295-2304
- Yufei Zhang, Jeffrey O. Kephart, Zijun Cui, Qiang Ji:
PhysPT: Physics-aware Pretrained Transformer for Estimating Human Dynamics from Monocular Videos. 2305-2317
- Arnab Kumar Mondal, Stefano Alletto, Denis Tomè:
HumMUSS: Human Motion Understanding Using State Space Models. 2318-2330
- Nicolas Ugrinovic, Boxiao Pan, Georgios Pavlakos, Despoina Paschalidou, Bokui Shen, Jordi Sanchez-Riera, Francesc Moreno-Noguer, Leonidas J. Guibas:
MultiPhys: Multi-Person Physics-Aware 3D Motion Estimation. 2331-2340
- Haowen Luo, Yunze Liu, Li Yi:
Physics-Aware Hand-Object Interaction Denoising. 2341-2350
- Supreeth Narasimhaswamy, Huy Anh Nguyen, Lihan Huang, Minh Hoai:
HOIST-Former: Hand-Held Objects Identification, Segmentation, and Tracking in the Wild. 2351-2361
- Soubhik Sanyal, Partha Ghosh, Jinlong Yang, Michael J. Black, Justus Thies, Timo Bolkart:
SCULPT: Shape-Conditioned Unpaired Learning of Pose-dependent Clothed and Textured Human Meshes. 2362-2371
- Tuomas Varanka, Tapani Toivonen, Soumya Tripathy, Guoying Zhao, Erman Acar:
PFStorer: Personalized Face Restoration and Super-Resolution. 2372-2381
- Pengfei Xie, Wenqiang Xu, Tutian Tang, Zhenjun Yu, Cewu Lu:
MS-MANO: Enabling Hand Pose Tracking with Biomechanical Constraints. 2382-2392
- Wenqian Zhang, Molin Huang, Yuxuan Zhou, Juze Zhang, Jingyi Yu, Jingya Wang, Lan Xu:
BOTH2Hands: Inferring 3D Hands from Both Text Prompts and Body Dynamics. 2393-2404
- Eric-Tuan Le, Antonis Kakolyris, Petros Koutras, Himmy Tam, Efstratios Skordos, George Papandreou, Riza Alp Güler, Iasonas Kokkinos:
MeshPose: Unifying DensePose and 3D Body Mesh reconstruction. 2405-2414
- Xi Liu, Ying Guo, Cheng Zhen, Tong Li, Yingying Ao, Pengfei Yan:
CustomListener: Text-Guided Responsive Interaction for User-Friendly Listening Head Generation. 2415-2424
- Jiayi Liang, Haotian Liu, Hongteng Xu, Dixin Luo:
Generalizable Face Landmarking Guided by Conditional Face Warping. 2425-2435
- Xinshun Wang, Zhongbin Fang, Xia Li, Xiangtai Li, Chen Chen, Mengyuan Liu:
Skeleton-in-Context: Unified Skeleton Sequence Modeling with In-Context Learning. 2436-2446
- Reni Paskaleva, Mykyta Holubakha, Andela Ilic, Saman Motamed, Luc Van Gool, Danda Pani Paudel:
A Unified and Interpretable Emotion Representation and Expression Generation. 2447-2456
- Yingyan Xu, Prashanth Chandran, Sebastian Weiss, Markus Gross, Gaspard Zoss, Derek Bradley:
Artist-Friendly Relightable and Animatable Neural Heads. 2457-2467
- Supreeth Narasimhaswamy, Uttaran Bhattacharya, Xiang Chen, Ishita Dasgupta, Saayan Mitra, Minh Hoai:
HandDiffuser: Text-to-Image Generation with Realistic Hand Appearances. 2468-2479
- Abhishek Tandon, Anujraaj Goyal, Henry M. Clever, Zackory Erickson:
BodyMAP - Jointly Predicting Body Mesh and 3D Applied Pressure Map for People in Bed. 2480-2489
- George Retsinas, Panagiotis Paraskevas Filntisis, Radek Danecek, Victoria Fernández Abrevaya, Anastasios Roussos, Timo Bolkart, Petros Maragos:
3D Facial Expressions through Analysis-by-Neural-Synthesis. 2490-2501
- Vinkle Srivastav, Keqi Chen, Nicolas Padoy:
SelfPose3d: Self-Supervised Multi-Person Multi-View 3d Pose Estimation. 2502-2512
- Tom Van Wouwe, Seunghwan Lee, Antoine Falisse, Scott L. Delp, C. Karen Liu:
DiffusionPoser: Real-Time Human Motion Reconstruction From Arbitrary Sparse Sensors Using Autoregressive Diffusion. 2513-2523
- Tian Ye, Sixiang Chen, Wenhao Chai, Zhaohu Xing, Jing Qin, Ge Lin, Lei Zhu:
Learning Diffusion Texture Priors for Image Restoration. 2524-2534
- Shangchen Zhou, Peiqing Yang, Jianyi Wang, Yihang Luo, Chen Change Loy:
Upscale-A-Video: Temporal-Consistent Diffusion Model for Real-World Video Super-Resolution. 2535-2545
- Kai Xu, Ziwei Yu, Xin Wang, Michael Bi Mi, Angela Yao:
Enhancing Video Super-Resolution via Implicit Resampling-based Alignment. 2546-2555
- Xinjie Zhang, Ren Yang, Dailan He, Xingtong Ge, Tongda Xu, Yan Wang, Hongwei Qin, Jun Zhang:
Boosting Neural Representations for Videos with a Conditional Decoder. 2556-2566
- Zheng Ding, Xuaner Zhang, Zhuowen Tu, Zhihao Xia:
Restoration by Generation with Constrained Priors. 2567-2577



- Pingping Zhang, Tianyu Yan, Yang Liu, Huchuan Lu:
Fantastic Animals and Where to Find Them: Segment Any Marine Animal with Dual SAM. 2578-2587
- Shay Dekel, Yosi Keller, Martin Cadík:
Estimating Extreme 3D Image Rotations using Cascaded Attention. 2588-2598
- Kanglong Fan, Wen Wen, Mu Li, Yifan Peng, Kede Ma:
Learned Scanpaths Aid Blind Panoramic Video Quality Assessment. 2599-2608
- Xiaoyan Cong, Yue Wu, Qifeng Chen, Chenyang Lei:
Automatic Controllable Colorization via Imagination. 2609-2619
- Chenxi Qiu, Tao Yue, Xuemei Hu:
Reconstruction-free Cascaded Adaptive Compressive Sensing. 2620-2630
- Xiaofeng Cong, Jie Gui, Jing Zhang, Junming Hou, Hao Shen:
A Semi-Supervised Nighttime Dehazing Baseline with Spatial-Frequency Aware and Realistic Brightness Constraint. 2631-2640
- Cheeun Hong, Kyoung Mu Lee:
AdaBM: On-the-Fly Adaptive Bit Mapping for Image Super-Resolution. 2641-2650
- Jaeha Kim, Junghun Oh, Kyoung Mu Lee:
Beyond Image Super-Resolution for Image Recognition with Task-Driven Perceptual Loss. 2651-2661
- Kangmin Xu, Liang Liao, Jing Xiao, Chaofeng Chen, Haoning Wu, Qiong Yan, Weisi Lin:
Boosting Image Quality Assessment Through Efficient Transformer Adaptation with Local Feature Enhancement. 2662-2672
- Guilherme A. Potje, Felipe Cadar, André Araújo, Renato Martins, Erickson R. Nascimento:
XFeat: Accelerated Features for Lightweight Image Matching. 2682-2691
- Tianhao Zhou, Haipeng Li, Ziyi Wang, Ao Luo, Chen-Lin Zhang, Jiajun Li, Bing Zeng, Shuaicheng Liu:
RecDiffusion: Rectangling for Image Stitching with Diffusion Models. 2692-2701
- Xin Tian, Ke Xu, Rynson W. H. Lau:
Unsupervised Salient Instance Detection. 2702-2712
- Zhen Liu, Hao Zhu, Qi Zhang, Jingde Fu, Weibing Deng, Zhan Ma, Yanwen Guo, Xun Cao:
FINER: Flexible Spectral-Bias Tuning in Implicit NEural Representation by Variableperiodic Activation Functions. 2713-2722
- Donghun Ryou, Inju Ha, Hyewon Yoo, Dongwan Kim, Bohyung Han:
Robust Image Denoising Through Adversarial Frequency Mixup. 2723-2732
- Xin Gao, Tianheng Qiu, Xinyu Zhang, Hanlin Bai, Kang Liu, Xuan Huang, Hu Wei, Guoqing Zhang, Huaping Liu:
Efficient Multi-Scale Network with Learnable Discrete Wavelet Transform for Blind Motion Deblurring. 2733-2742
- Zhongyu Li, Lei Zhang:
Efficient Scene Recovery Using Luminous Flux Prior. 2743-2752
- Guangyang Wu, Xin Tao, Changlin Li, Wenyi Wang, Xiaohong Liu, Qingqing Zheng:
Perception-Oriented Video Frame Interpolation via Asymmetric Blending. 2753-2762
- Wen Wen, Mu Li, Yabin Zhang, Yiting Liao, Junlin Li, Li Zhang, Kede Ma:
Modular Blind Video Quality Assessment. 2763-2772
- Jiawei Liu, Qiang Wang, Huijie Fan, Yinong Wang, Yandong Tang, Liangqiong Qu:
Residual Denoising Diffusion Models. 2773-2783
- Woo Kyoung Han, Sunghoon Im, Jaedeok Kim, Kyong Hwan Jin:
JDEC: JPEG Decoding via Enhanced Continuous Cosine Coefficients. 2784-2793
- Agneet Chatterjee, Tejas Gokhale, Chitta Baral, Yezhou Yang:
On the Robustness of Language Guidance for Low-Level Vision Tasks: Findings from Depth Estimation. 2794-2803
- Bang-Dang Pham, Phong Tran, Anh Tuan Tran, Cuong Pham, Rang Nguyen, Minh Hoai:
Blur2Blur: Blur Conversion for Unsupervised Image Deblurring on Unknown Domains. 2804-2813
- Shiyan Chen, Jiyuan Zhang, Zhaofei Yu, Tiejun Huang:
Exploring Efficient Asymmetric Blind-Spots for Self-Supervised Denoising in Real-World Scenarios. 2814-2823
- Jiezhang Cao, Yue Shi, Kai Zhang, Yulun Zhang, Radu Timofte, Luc Van Gool:
Deep Equilibrium Diffusion Restoration with Parallel Sampling. 2824-2834
- Kun Yuan, Hongbo Liu, Mading Li, Muyi Sun, Ming Sun, Jiachao Gong, Jinhua Hao, Chao Zhou, Yansong Tang:
PTM-VQA: Efficient Video Quality Assessment Leveraging Diverse PreTrained Models from the Wild. 2835-2845
- Yafei Zhang, Shen Zhou, Huafeng Li:
Depth Information Assisted Collaborative Mutual Promotion Network for Single Image Dehazing. 2846-2855
- Leheng Zhang, Yawei Li, Xingyu Zhou, Xiaorui Zhao, Shuhang Gu:
Transcending the Limit of Local Window: Advanced Super-Resolution Transformer with Adaptive Token Dictionary. 2856-2865
- Jingbo Lin, Zhilu Zhang, Yuxiang Wei, Dongwei Ren, Dongsheng Jiang, Qi Tian, Wangmeng Zuo:
Improving Image Restoration Through Removing Degradations in Textual Representations. 2866-2878



- Yong Shu, Liquan Shen, Xiangyu Hu, Mengyao Li, Zihao Zhou:
Towards Real-World HDR Video Reconstruction: A Large-Scale Benchmark Dataset and A Two-Stage Alignment Network. 2879-2888
- Xingguang Zhang, Nicholas Chimitt, Yiheng Chi, Zhiyuan Mao, Stanley H. Chan:
Spatio-Temporal Turbulence Mitigation: A Translational Perspective. 2889-2899
- Xiaogang Xu, Shu Kong, Tao Hu, Zhe Liu, Hujun Bao:
Boosting Image Restoration via Priors from Pre-Trained Models. 2900-2909
- Zhangkai Ni, Juncheng Wu, Zian Wang, Wenhan Yang, Hanli Wang, Lin Ma:
Misalignment-Robust Frequency Distribution Loss for Image Transformation. 2910-2919
- Enxuan Gu, Hongwei Ge, Yong Guo:
CoDe: An Explicit Content Decoupling Framework for Image Restoration. 2920-2930
- Wei-Ting Chen, Gurunandan Krishnan, Qiang Gao, Sy-Yen Kuo, Sizhuo Ma, Jian Wang:
DSL-FIQA: Assessing Facial Image Quality via Dual-Set Degradation Learning and Landmark-Guided Transformer. 2931-2941
- Hyeongmin Lee, Kyoungkook Kang, Jungseul Ok, Sunghyun Cho:
CLIPtone: Unsupervised Learning for Text-Based Image Tone Adjustment. 2942-2951
- Shihao Zhou, Duosheng Chen, Jinshan Pan, Jinglei Shi, Jufeng Yang:
Adapt or Perish: Adaptive Sparse Transformer with Attentive Feature Refinement for Image Restoration. 2952-2963
- Qiang Zhu, Jinhua Hao, Yukang Ding, Yu Liu, Qiao Mo, Ming Sun, Chao Zhou, Shuyuan Zhu:
CPGA: Coding Priors-Guided Aggregation Network for Compressed Video Quality Enhancement. 2964-2974
- Kyunghyun Lee, Ukcheol Shin, Byeong-Uk Lee:
Learning to Control Camera Exposure via Reinforcement Learning. 2975-2983
- Ziwen Li, Feng Zhang, Meng Cao, Jinpu Zhang, Yuanjie Shao, Yuehuan Wang, Nong Sang:
Real-Time Exposure Correction via Collaborative Transformations and Adaptive Sampling. 2984-2994
- Jun Xiao, Zihang Lyu, Cong Zhang, Yakun Ju, Changjian Shui, Kin-Man Lam:
Towards Progressive Multi-Frequency Representation for Image Warping. 2995-3004
- Li Pang, Xiangyu Rui, Long Cui, Hongzhong Wang, Deyu Meng, Xiangyong Cao:
HIR-Diff: Unsupervised Hyperspectral Image Restoration Via Improved Diffusion Models. 3005-3014
- Yiqi Shi, Duo Liu, Liguo Zhang, Ye Tian, Xuezhi Xia, Xiaojing Fu:
ZERO-IG: Zero-Shot Illumination-Guided Joint Denoising and Adaptive Enhancement for Low-Light Images. 3015-3024
- Hamadi Chihaoui, Paolo Favaro:
Masked and Shuffled Blind Spot Denoising for Real-World Images. 3025-3034
- Huiyuan Fu, Fei Peng, Xianwei Li, Yefan Li, Xin Wang, Huadong Ma:
Continuous Optical Zooming: A Benchmark for Arbitrary-Scale Image Super-Resolution in Real World. 3035-3044
- Atefeh Khoshkhahtinat, Ali Zafari, Piyush M. Mehta, Nasser M. Nasrabadi:
Laplacian-guided Entropy Model in Neural Codec with Blur-dissipated Synthesis. 3045-3054
- Yuan Gao, Yuqing Zhu, Xinjun Li, Yimin Du, Tianzhu Zhang:
SD2Event: Self-Supervised Learning of Dynamic Detectors and Contextual Descriptors for Event Cameras. 3055-3064
- Lanyun Zhu, Tianrun Chen, Deyi Ji, Jieping Ye, Jun Liu:
LLaFS: When Large Language Models Meet Few-Shot Segmentation. 3065-3075
- Junyi Zhang, Charles Herrmann, Junhwa Hur, Eric Chen, Varun Jampani, Deqing Sun, Ming-Hsuan Yang:
Telling Left from Right: Identifying Geometry-Aware Semantic Correspondence. 3076-3085
- Gen Li, Deqing Sun, Laura Sevilla-Lara, Varun Jampani:
One-Shot Open Affordance Learning with Foundation Models. 3086-3096
- Boyuan Sun, Yuqi Yang, Le Zhang, Ming-Ming Cheng, Qibin Hou:
CorrMatch: Label Propagation via Correlation Matching for Semi-Supervised Semantic Segmentation. 3097-3107
- Yasser Benigmim, Subhankar Roy, Slim Essid, Vicky Kalogeiton, Stéphane Lathuilière:
Collaborating Foundation Models for Domain Generalized Semantic Segmentation. 3108-3119
- You Huang, Zongyu Lan, Liujuan Cao, Xianming Lin, Shengchuan Zhang, Guannan Jiang, Rongrong Ji:
FocSAM: Delving Deeply into Focused Objects in Segmenting Anything. 3120-3130
- Simon Weber, Thomas Dagès, Maolin Gao, Daniel Cremers:
Finsler-Laplace-Beltrami Operators with Application to Shape Analysis. 3131-3140
- Yijia Weng, Bowen Wen, Jonathan Tremblay, Valts Blukis, Dieter Fox, Leonidas J. Guibas, Stan Birchfield:
Neural Implicit Representation for Building Digital Twins of Unknown Articulated Objects. 3141-3150
- Ho Kei Cheng, Seoung Wug Oh, Brian L. Price, Joon-Young Lee, Alexander G. Schwing:
Putting the Object Back into Video Object Segmentation. 3151-3161
- Yiran Song, Qianyu Zhou, Xiangtai Li, Deng-Ping Fan, Xuequan Lu, Lizhuang Ma:
BA-SAM: Scalable Bias-Mode Attention Mask for Segment Anything Model. 3162-3173



- Daan de Geus, Gijs Dubbelman:
Task-Aligned Part-Aware Panoptic Segmentation Through Joint Object-Part Representations. 3174-3183
- Matteo Sodano, Federico Magistri, Lucas Nunes, Jens Behley, Cyrill Stachniss:
Open-World Semantic Segmentation Including Class Similarity. 3184-3194
- Thomas V. Chang, Simon Seibt, Bartosz von Rymon Lipinski:
Hierarchical Histogram Threshold Segmentation - Auto-terminating High-detail Oversegmentation. 3195-3204
- Duojun Huang, Xinyu Xiong, Jie Ma, Jichang Li, Zequn Jie, Lin Ma, Guanbin Li:
AlignSAM: Aligning Segment Anything Model to Open Context via Reinforcement Learning. 3205-3215
- Yichen Liu, Benran Hu, Chi-Keung Tang, Yu-Wing Tai:
SANeRF-HQ: Segment Anything for NeRF in High Quality. 3216-3226
- Minghan Li, Shuai Li, Xindong Zhang, Lei Zhang:
UniVS: Unified and Universal Video Segmentation with Prompts as Queries. 3227-3238
- Bedrettin Cetinkaya, Sinan Kalkan, Emre Akbas:
RankED: Addressing Imbalance and Uncertainty in Edge Detection Using Ranking-based Losses. 3239-3249
- Hebei Li, Jin Wang, Jiahui Yuan, Yue Li, Wenming Weng, Yansong Peng, Yueyi Zhang, Zhiwei Xiong, Xiaoyan Sun:
Event-Assisted Low-Light Video Object Segmentation. 3250-3259
- Jianan Li, Qiulei Dong:
Density-Guided Semi-Supervised 3D Semantic Segmentation with Dual-Space Hardness Sampling. 3260-3269
- Yi Zhang, Meng-Hao Guo, Miao Wang, Shi-Min Hu:
Exploring Regional Clues in CLIP for Zero-Shot Semantic Segmentation. 3270-3280
- Yichen Li, Kaichun Mo, Yueqi Duan, He Wang, Jiequan Zhang, Lin Shao, Wojciech Matusik, Leonidas J. Guibas:
Category-Level Multi-Part Multi-Joint 3D Shape Assembly. 3281-3291
- Yingda Yin, Yuzheng Liu, Yang Xiao, Daniel Cohen-Or, Jingwei Huang, Baoquan Chen:
SAI3D: Segment any Instance in 3D Scenes. 3292-3302
- Xiaoyang Wang, Huihui Bai, Limin Yu, Yao Zhao, Jimin Xiao:
Towards the Uncharted: Density-Descending Feature Perturbation for Semi-supervised Semantic Segmentation. 3303-3312
- Lennart Bastian, Yizheng Xie, Nassir Navab, Zorah Lähner:
Hybrid Functional Maps for Crease-Aware Non-Isometric Shape Matching. 3313-3323
- Feilong Tang, Zhongxing Xu, Zhaojun Qu, Wei Feng, Xingjian Jiang, Zongyuan Ge:
Hunting Attributes: Context Prototype-Aware Learning for Weakly Supervised Semantic Segmentation. 3324-3334
- Jiawei Liu, Changkun Ye, Ruikai Cui, Nick Barnes:
Self-Calibrating Vicinal Risk Minimisation for Model Calibration. 3335-3345
- Beomyoung Kim, Joonsang Yu, Sung Ju Hwang:
ECLIPSE: Efficient Continual Learning in Panoptic Segmentation with Visual Prompt Tuning. 3346-3356
- Yuhang Ding, Liulei Li, Wenguan Wang, Yi Yang:
Clustering Propagation for Universal Medical Image Segmentation. 3357-3369
- Lanyun Zhu, Tianrun Chen, Jianxiong Yin, Simon See, Jun Liu:
Addressing Background Context Bias in Few-Shot Segmentation Through Iterative Modulation. 3370-3379
- Jiahao Nie, Yun Xing, Gongjie Zhang, Pei Yan, Aoran Xiao, Yap-Peng Tan, Alex C. Kot, Shijian Lu:
Cross-Domain Few-Shot Segmentation via Iterative Support-Query Correspondence Mining. 3380-3390
- Huayu Mai, Rui Sun, Tianzhu Zhang, Feng Wu:
RankMatch: Exploring the Better Consistency Regularization for Semi-Supervised Semantic Segmentation. 3391-3401
- Xiang Li, Jinglu Wang, Xiaohao Xu, Xiulian Peng, Rita Singh, Yan Lu, Bhiksha Raj:
QDFomer: Towards Robust Audiovisual Segmentation in Complex Environments with Quantization-based Semantic Decomposition. 3402-3413
- Linwei Chen, Lin Gu, Dezhi Zheng, Ying Fu:
Frequency-Adaptive Dilated Convolution for Semantic Segmentation. 3414-3425
- Bin Xie, Jiale Cao, Jin Xie, Fahad Shahbaz Khan, Yanwei Pang:
SED: A Simple Encoder-Decoder for Open-Vocabulary Semantic Segmentation. 3426-3436
- Xinqiao Zhao, Ziqian Yang, Tianhong Dai, Bingfeng Zhang, Jimin Xiao:
PSDPM: Prototype-based Secondary Discriminative Pixels Mining for Weakly Supervised Semantic Segmentation. 3437-3446
- Matteo Bastico, Etienne Decencière, Laurent Corté, Yannick Tillier, David Ryckelynck:
Coupled Laplacian Eigenmaps for Locally-Aware 3D Rigid Point Cloud Matching. 3447-3458
- Yong Liu, Cairong Zhang, Yitong Wang, Jiahao Wang, Yujiu Yang, Yansong Tang:
Universal Segmentation at Arbitrary Granularity with Language Instruction. 3459-3469
- Ardian Umam, Cheng-Kun Yang, Min-Hung Chen, Jen-Hui Chuang, Yen-Yu Lin:
PartDistill: 3D Shape Part Segmentation by Vision-Language Model Distillation. 3470-3479



- Marilyn Keller, Vaibhav Arora, Abdelmouttaleb Dakri, Shivam Chandhok, Jürgen Machann, Andreas Fritzsche, Michael J. Black, Sergi Pujades:
HIT: Estimating Internal Human Implicit Tissues from the Body Surface. 3480-3490
- Yong Liu, Sule Bai, Guanbin Li, Yitong Wang, Yansong Tang:
Open-Vocabulary Segmentation with Semantic-Assisted Calibration. 3491-3500
- Yian Zhao, Kehan Li, Zesen Cheng, Pengchong Qiao, Xiawu Zheng, Rongrong Ji, Chang Liu, Li Yuan, Jie Chen:
GraCo: Granularity-Controllable Interactive Segmentation. 3501-3510
- Zhiheng Cheng, Qingyue Wei, Hongru Zhu, Yan Wang, Liangqiong Qu, Wei Shao, Yuyin Zhou:
Unleashing the Potential of SAM for Medical Adaptation via Hierarchical Decoding. 3511-3522
- Chanyoung Kim, Woojung Han, Dayun Ju, Seong Jae Hwang:
EAGLE: Eigen Aggregation Learning for Object-Centric Unsupervised Semantic Segmentation. 3523-3533
- Yuanchen Wu, Xichen Ye, Kequan Yang, Jide Li, Xiaoqiang Li:
DuPL: Dual Student with Trustworthy Progressive Learning for Robust Weakly Supervised Semantic Segmentation. 3534-3543
- Diandian Guo, Deng-Ping Fan, Tongyu Lu, Christos Sakaridis, Luc Van Gool:
Vanishing-Point-Guided Video Semantic Segmentation of Driving Scenes. 3544-3553
- Junjiao Tian, Lavisha Aggarwal, Andrea Colaco, Zsolt Kira, Mar González-Franco:
Diffuse, Attend, and Segment: Unsupervised Zero-Shot Segmentation using Stable Diffusion. 3554-3563
- Ayush Jain, Pushkal Katara, Nikolaos Gkanatsios, Adam W. Harley, Gabriel Sarch, Kriti Aggarwal, Vishrav Chaudhary, Katerina Fragkiadaki:
ODIN: A Single Model for 2D and 3D Segmentation. 3564-3574
- Jiafan Zhuang, Zilei Wang, Yixin Zhang, Zhun Fan:
Infer from What You Have Seen Before: Temporally-dependent Classifier for Semi-supervised Video Segmentation. 3575-3584
- Zhaoyang Wei, Pengfei Chen, Xuehui Yu, Guorong Li, Jianbin Jiao, Zhenjun Han:
Semantic-aware SAM for Point-Prompted Instance Segmentation. 3585-3594
- Sung-Hoon Yoon, Hoyong Kwon, Hyeonseong Kim, Kuk-Jin Yoon:
Class Tokens Infusion for Weakly Supervised Semantic Segmentation. 3595-3605
- Zhiwei Yang, Kexue Fu, Minghong Duan, Linhao Qu, Shuo Wang, Zhijian Song:
Separate and Conquer: Decoupling Co-occurrence via Decomposition and Representation for Weakly Supervised Semantic Segmentation. 3606-3615
- Huicong Zhang, Haozhe Xie, Hongxun Yao:
Blur-Aware Spatio-Temporal Sparse Transformer for Video Deblurring. 3616-3626
- Woo-Jin Ahn, Geun-Yeong Yang, Hyun Duck Choi, Myo-Taeg Lim:
Style Blind Domain Generalized Semantic Segmentation via Covariance Alignment and Semantic Consistence Contrastive Learning. 3616-3626
- Haonan Wang, Qixiang Zhang, Yi Li, Xiaomeng Li:
AllSpark: Reborn Labeled Features from Unlabeled in Transformer for Semi-Supervised Semantic Segmentation. 3627-3636
- Leon Sick, Dominik Engel, Pedro Hermosilla, Timo Ropinski:
Unsupervised Semantic Segmentation Through Depth-Guided Feature Correlation and Sampling. 3637-3646
- Nissim Maruani, Maks Ovsjanikov, Pierre Alliez, Mathieu Desbrun:
PoNQ: A Neural QEM-Based Mesh Representation. 3647-3657
- Dongliang Cao, Marvin Eisenberger, Nafie El Amrani, Daniel Cremers, Florian Bernard:
Spectral Meets Spatial: Harmonising 3D Shape Matching and Interpolation. 3658-3668
- Jiayi Zhu, Qing Guo, Felix Juefei-Xu, Yihao Huang, Yang Liu, Geguang Pu:
Cosalpure: Learning Concept from Group Images for Robust Co-Saliency Detection. 3669-3678
- Jiawei Wang, Changjian Li:
ContextSeg: Sketch Semantic Segmentation by Querying the Context with Attention. 3679-3688
- Luca Barsellotti, Roberto Amoroso, Marcella Cornia, Lorenzo Baraldi, Rita Cucchiara:
Training-Free Open-Vocabulary Segmentation with Offline Diffusion-Augmented Prototype Generation. 3689-3698
- Bo Li, Haoke Xiao, Lv Tang:
ASAM: Boosting Segment Anything Model with Adversarial Tuning. 3699-3710
- He Guo, Zixuan Ye, Zhiguo Cao, Hao Lu:
In-Context Matting. 3711-3720
- Hyeokjun Kweon, Jihun Kim, Kuk-Jin Yoon:
Weakly Supervised Point Cloud Semantic Segmentation via Artificial Oracle. 3721-3731
- Changki Sung, Wanhee Kim, Jungho An, Wooju Lee, Hyungtae Lim, Hyun Myung:
ContextContrast: Contextual Contrastive Learning for Semantic Segmentation. 3732-3742
- Zelin Peng, Zhengqin Xu, Zhilin Zeng, Lingxi Xie, Qi Tian, Wei Shen:
Parameter Efficient Fine-Tuning via Cross Block Orchestration for Segment Anything Model. 3743-3752



- Haocheng Yuan, Jing Xu, Hao Pan, Adrien Bousseau, Niloy J. Mitra, Changjian Li:
CADTalk: An Algorithm and Benchmark for Semantic Commenting of CAD Programs. 3753-3762
- Yujia Liu, Anton Obukhov, Jan Dirk Wegner, Konrad Schindler:
Point2CAD: Reverse Engineering CAD Models from 3D Point Clouds. 3763-3772
- Qin Liu, Jaemin Cho, Mohit Bansal, Marc Niethammer:
Rethinking Interactive Image Segmentation with Low Latency, High Quality, and Diverse Prompts. 3773-3782
- Junfeng Wu, Yi Jiang, Qihao Liu, Zehuan Yuan, Xiang Bai, Song Bai:
General Object Foundation Model for Images and Videos at Scale. 3783-3795
- Bingfeng Zhang, Siyue Yu, Yunchao Wei, Yao Zhao, Jimin Xiao:
Frozen CLIP: A Strong Backbone for Weakly Supervised Semantic Segmentation. 3796-3806
- Minhyeok Lee, Suhwan Cho, Dogyoon Lee, Chaewon Park, Jungho Lee, Sangyoun Lee:
Guided Slot Attention for Unsupervised Video Object Segmentation. 3807-3816
- Ziqin Zhou, Hai-Ming Xu, Yangyang Shu, Lingqiao Liu:
Unlocking the Potential of Pre-Trained Vision Transformers for Few-Shot Semantic Segmentation through Relationship Descriptors. 3817-3827
- Walid Bousselham, Felix Petersen, Vittorio Ferrari, Hilde Kuehne:
Grounding Everything: Emerging Localization Properties in Vision-Language Transformers. 3828-3837
- Xiangyang Zhu, Renrui Zhang, Bowei He, Ziyu Guo, Jiaming Liu, Han Xiao, Chaoyou Fu, Hao Dong, Peng Gao:
No Time to Train: Empowering Non-Parametric Networks for Few-Shot 3D Scene Segmentation. 3838-3847
- Yizheng Gong, Siyue Yu, Xiaoyang Wang, Jimin Xiao:
Continual Segmentation with Disentangled Objectness Learning and Class Recognition. 3848-3857
- Zhuofan Xia, Dongchen Han, Yizeng Han, Xuran Pan, Shiji Song, Gao Huang:
GSVA: Generalized Segmentation via Multimodal Large Language Models. 3858-3869
- Chuong Huynh, Seoung Wug Oh, Abhinav Shrivastava, Joon-Young Lee:
MaGGle: Masked Guided Gradual Human Instance Matting. 3870-3879
- Zitao Wang, Qiguang Miao, Yue Xi, Peipei Zhao:
EFormer: Enhanced Transformer Towards Semantic-Contour Features of Foreground for Portraits Matting. 3880-3889
- Zhiwen Chen, Zhiyu Zhu, Yifan Zhang, Junhui Hou, Guangming Shi, Jinjian Wu:
Segment Any Event Streams via Weighted Adaptation of Pivotal Tokens. 3890-3900
- Kenji Enomoto, TJ Rhodes, Brian Price, Gavin Miller:
Polar Matte: Fully Computational Ground-Truth-Quality Alpha Matte Extraction for Images and Video using Polarized Screen Matting. 3901-3909
- Wenjie Zhao, Jia Li, Xin Dong, Yu Xiang, Yunhui Guo:
Segment Every Out-of-Distribution Object. 3910-3920
- Qian Yu, Xiaoqi Zhao, Youwei Pang, Lihe Zhang, Huchuan Lu:
Multi-View Aggregation Network for Dichotomous Image Segmentation. 3921-3930
- Ege Ozguroglu, Ruoshi Liu, Dídac Surís, Dian Chen, Achal Dave, Pavel Tokmakov, Carl Vondrick:
pix2gestalt: Amodal Segmentation by Synthesizing Wholes. 3931-3940
- Jin Wang, Bingfeng Zhang, Jian Pang, Honglong Chen, Weifeng Liu:
Rethinking Prior Information Generation with CLIP for Few-Shot Segmentation. 3941-3951
- Yuan Wang, Rui Sun, Naisong Luo, Yuwen Pan, Tianzhu Zhang:
Image-to-Image Matching via Foundation Models: A New Perspective for Open-Vocabulary Semantic Segmentation. 3952-3963
- Zijian Wu, Jun Lu, Jing Han, Lianfa Bai, Yi Zhang, Zhuang Zhao, Siyang Song:
Domain Separation Graph Neural Networks for Saliency Object Ranking. 3964-3974
- Sandra Kara, Hejer Ammar, Julien Denize, Florian Chabot, Quoc-Cuong Pham:
DIOD: Self-Distillation Meets Object Discovery. 3975-3985
- Chengxiang Fan, Muzhi Zhu, Hao Chen, Yang Liu, Weijia Wu, Huaiqi Zhang, Chunhua Shen:
DiverGen: Improving Instance Segmentation by Learning Wider Data Distribution with More Diverse Generative Data. 3986-3995
- Zhaochong An, Guolei Sun, Yun Liu, Fayao Liu, Zongwei Wu, Dan Wang, Luc Van Gool, Serge J. Belongie:
Rethinking Few-shot 3D Point Cloud Semantic Segmentation. 3996-4006
- Xinting Hu, Li Jiang, Bernt Schiele:
Training Vision Transformers for Semi-Supervised Semantic Segmentation. 4007-4017
- Phuc D. A. Nguyen, Tuan Duc Ngo, Evangelos Kalogerakis, Chuang Gan, Anh Tuan Tran, Cuong Pham, Khoi Nguyen:
Open3DIS: Open-Vocabulary 3D Instance Segmentation with 2D Mask Guidance. 4018-4028
- Jiayun Luo, Siddhesh Khandelwal, Leonid Sigal, Boyang Li:
Emergent Open-Vocabulary Semantic Segmentation from Off-the-Shelf Vision-Language Models. 4029-4040
- Robin Magnet, Maks Ovsjanikov:
Memory-Scalable and Simplified Functional Map Learning. 4041-4050



- Chaewon Lee, Seon-Ho Lee, Chang-Su Kim:
MFP: Making Full Use of Probability Maps for Interactive Image Segmentation. 4051-4059
- Sangyun Shin, Kaichen Zhou, Madhu Vankadari, Andrew Markham, Niki Trigoni:
Spherical Mask: Coarse-to-Fine 3D Point Cloud Instance Segmentation with Spherical Representation. 4060-4069
- Hanyang Chi, Jian Pang, Bingfeng Zhang, Weifeng Liu:
Adaptive Bidirectional Displacement for Semi-Supervised Medical Image Segmentation. 4070-4080
- Wei-Ting Chen, Yu-Jiet Vong, Sy-Yen Kuo, Sizhuo Ma, Jian Wang:
RobustSAM: Segment Anything Robustly on Degraded Images. 4081-4091
- Pancheng Zhao, Peng Xu, Pengda Qin, Deng-Ping Fan, Zhicheng Zhang, Guoli Jia, Bowen Zhou, Jufeng Yang:
LAKE-RED: Camouflaged Images Generation by Latent Background Knowledge Retrieval-Augmented Diffusion. 4092-4101
- Jingyun Wang, Guoliang Kang:
Learn to Rectify the Bias of CLIP for Unsupervised Semantic Segmentation. 4102-4112
- Seokju Cho, Heeseong Shin, Sunghwan Hong, Anurag Arnab, Paul Hongsuck Seo, Seungryong Kim:
CAT-Seg: Cost Aggregation for Open-Vocabulary Semantic Segmentation. 4113-4123
- Chao Shang, Zichen Song, Heqian Qiu, Lanxiao Wang, Fanman Meng, Hongliang Li:
Prompt-Driven Referring Image Segmentation with Instance Contrasting. 4124-4134
- Joren Brunekreef, Eric Marcus, Ray Sheombarsing, Jan-Jakob Sonke, Jonas Teuwen:
Kandinsky Conformal Prediction: Efficient Calibration of Image Segmentation Algorithms. 4135-4143
- Xiongwei Wu, Sicheng Yu, Ee-Peng Lim, Chong-Wah Ngo:
OVFoodSeg: Elevating Open-Vocabulary Food Image Segmentation via Image-Informed Textual Representation. 4144-4153
- Thomas Wimmer, Peter Wonka, Maks Ovsjanikov:
Back to 3D: Few-Shot 3D Keypoint Detection with Back-Projected 2D Features. 4154-4164
- Xiao Zhang, David Yunis, Michael Maire:
Deciphering 'What' and 'Where' Visual Pathways from Spectral Clustering of Layer-Distributed Neural Representations. 4165-4175
- Ahmed Bourouis, Judith Ellen Fan, Yulia Gryaditskaya:
Open Vocabulary Semantic Scene Sketch Understanding. 4176-4186
- Xiaoqi Wang, Wenbin He, Xiwei Xuan, Clint Sebastian, Jorge Piazentin Ono, Xin Li, Sima Behpour, Thang Doan, Liang Gou, Han-Wei Shen, Liu Ren:
USE: Universal Segment Embeddings for Open-Vocabulary Image Segmentation. 4187-4196
- Yuhao Liu, Zhanghan Ke, Fang Liu, Nanxuan Zhao, Rynson W. H. Lau:
Diff-Plugin: Revitalizing Details for Diffusion-Based Low-Level Tasks. 4197-4208
- Xuanchi Ren, Jiahui Huang, Xiaohui Zeng, Ken Museth, Sanja Fidler, Francis Williams:
XCube: Large-Scale 3D Generative Modeling using Sparse Voxel Hierarchies. 4209-4219
- Yi-Hua Huang, Yang-Tian Sun, Ziyi Yang, Xiaoyang Lyu, Yan-Pei Cao, Xiaojuan Qi:
SC-GS: Sparse-Controlled Gaussian Splatting for Editable Dynamic Scenes. 4220-4230
- Anand Bhattad, James Soole, David A. Forsyth:
StyLitGAN: Image-Based Relighting via Latent Control. 4231-4240
- Jiraphon Yenphraphai, Xichen Pan, Sainan Liu, Daniele Panozzo, Saining Xie:
Image Sculpting: Precise Object Editing with 3D Geometry Control. 4241-4251
- Xianfang Zeng, Xin Chen, Zhongqi Qi, Wen Liu, Zibo Zhao, Zhibin Wang, Bin Fu, Yong Liu, Gang Yu:
Paint3D: Paint Anything 3D With Lighting-Less Texture Diffusion Models. 4252-4262
- Yiqun Mei, Yu Zeng, He Zhang, Zhixin Shu, Xuaner Zhang, Sai Bi, Jianming Zhang, HyunJoon Jung, Vishal M. Patel:
Holo-Relighting: Controllable Volumetric Portrait Relighting from a Single Image. 4263-4273
- Daniel Rebain, Soroosh Yazdani, Kwang Moo Yi, Andrea Tagliasacchi:
Neural Fields as Distributions: Signal Processing Beyond Euclidean Space. 4274-4283
- Jialun Liu, Cheming Wu, Xinqi Liu, Xing Liu, Jinbo Wu, Haotian Peng, Chen Zhao, Haocheng Feng, Jingtuo Liu, Errui Ding:
TexOct: Generating Textures of 3D Models with Octree-based Diffusion. 4284-4293
- Yishun Dou, Zhong Zheng, Qiaoqiao Jin, Rui Shi, Yuhan Li, Bingbing Ni:
Differentiable Micro-Mesh Construction. 4294-4303
- Yu-Ying Yeh, Jia-Bin Huang, Changil Kim, Lei Xiao, Thu Nguyen-Phuoc, Numair Khan, Cheng Zhang, Manmohan Chandraker, Carl S. Marshall, Zhao Dong, Zhengqin Li:
TextureDreamer: Image-Guided Texture Synthesis through Geometry-Aware Diffusion. 4304-4314
- Seungwoo Yoo, Kunho Kim, Vladimir G. Kim, Minhyuk Sung:
As-Plausible-As-Possible: Plausibility-Aware Mesh Deformation Using 2D Diffusion Priors. 4315-4324
- Rinon Gal, Yael Vinker, Yuval Alaluf, Amit Bermano, Daniel Cohen-Or, Ariel Shamir, Gal Chechik:
Breathing Life Into Sketches Using Text-to-Video Priors. 4325-4336



- Yishun Dou, Zhong Zheng, Qiaoqiao Jin, Bingbing Ni, Yugang Chen, Junxiang Ke:
Real-Time Neural BRDF with Spherically Distributed Primitives. 4337-4346
- Kim Youwang, Tae-Hyun Oh, Gerard Pons-Moll:
Paint-it: Text-to-Texture Synthesis via Deep Convolutional Texture Map Optimization and Physically-Based Rendering. 4347-4356
- Jia Li, Ziling Chen, Xiaolong Wu, Lu Wang, Beibei Wang, Lei Zhang:
Neural Super-Resolution for Real-Time Rendering with Radiance Demodulation. 4357-4367
- Yifei Li, Hsiao-Yu Chen, Egor Larionov, Nikolaos Sarafianos, Wojciech Matusik, Tuur Stuyck:
DiffAvatar: Simulation-Ready Garment Optimization with Differentiable Simulation. 4368-4378
- Ivan Lopes, Fabio Pizzati, Raoul de Charette:
Material Palette: Extraction of Materials from a Single Image. 4379-4388
- Tianyi Xie, Zeshun Zong, Yuxing Qiu, Xuan Li, Yutao Feng, Yin Yang, Chenfanfu Jiang:
PhysGaussian: Physics-Integrated 3D Gaussians for Generative Dynamics. 4389-4398
- Hoon-Gyu Chung, Seokjun Choi, Seung-Hwan Baek:
Differentiable Point-Based Inverse Rendering. 4399-4408
- Justine Giroux, Mohammad Reza Karimi Dastjerdi, Yannick Hold-Geoffroy, Javier Vazquez-Corral, Jean-François Lalonde:
Towards a Perceptual Evaluation Framework for Lighting Estimation. 4410-4419
- Zhongyin Zhao, Ye Chen, Zhangli Hu, Xuanhong Chen, Bingbing Ni:
Vector Graphics Generation via Mutually Impulsed Dual-Domain Diffusion. 4420-4428
- Giuseppe Vecchio, Renato Sortino, Simone Palazzo, Concetto Spampinato:
MatFuse: Controllable Material Generation with Diffusion Models. 4429-4438
- Carlos Rodríguez-Pardo, Dan Casas, Elena Garces, Jorge Lopez-Moreno:
TexTile: A Differentiable Metric for Texture Tileability. 4439-4449
- Yutao Feng, Yintong Shang, Xuan Li, Tianjia Shao, Chenfanfu Jiang, Yin Yang:
PIE-NeRF: Physics-Based Interactive Elastodynamics with NeRF. 4450-4461
- Jiahao Ma, Miaomiao Liu, David Ahmedt-Aristizabal, Chuong Nguyen:
HashPoint: Accelerated Point Searching and Sampling for Neural Rendering. 4462-4472
- Dale Decatur, Itai Lang, Kfir Aberman, Rana Hanocka:
3D Paintbrush: Local Stylization of 3D Shapes with Cascaded Score Distillation. 4473-4483
- Miguel Fainstein, Viviana Siless, Emmanuel Iarussi:
DUDF: Differentiable Unsigned Distance Fields with Hyperbolic Scaling. 4484-4493
- Niladri Shekhar Dutt, Sanjeev Muralikrishnan, Niloy J. Mitra:
Diffusion 3D Features (Diff3F) Decorating Untextured Shapes with Distilled Semantic Features. 4494-4504
- Soyeon Yoon, Kwan Yun, Kwanggyoong Seo, Sihun Cha, Jung Eun Yoo, Junyoung Noh:
LeGO: Leveraging a Surface Deformation Network for Animatable Stylized Face Generation with One Example. 4505-4514
- Yichen Sheng, Zixun Yu, Lu Ling, Zhiwen Cao, Xuaner Zhang, Xin Lu, Ke Xian, Haiting Lin, Bedrich Benes:
Dr.Bokeh: DiffeRentiable Occlusion-Aware Bokeh Rendering. 4515-4525
- Xiaoliang Ju, Zhaoyang Huang, Yijiin Li, Guofeng Zhang, Yu Qiao, Hongsheng Li:
DiffInDScape: Diffusion-Based High-Quality 3D Indoor Scene Generation. 4526-4535
- Xuecan Wang, Shibang Xiao, Xiaohui Liang :
LightOctree: Lightweight 3D Spatially-Coherent Indoor Lighting Estimation. 4536-4545
- Ximing Xing, Haitao Zhou, Chuang Wang, Jing Zhang, Dong Xu, Qian Yu:
SVGDreamer: Text Guided SVG Generation with Diffusion Model. 4546-4555
- Ruizhi Shao, Jingxiang Sun, Cheng Peng, Zerong Zheng, Boyao Zhou, Hongwen Zhang, Yebin Liu:
Control4D: Efficient 4D Portrait Editing With Text. 4556-4567
- Xin Huang, Ruizhi Shao, Qi Zhang, Hongwen Zhang, Ying Feng, Yebin Liu, Qing Wang:
HumanNorm: Learning Normal Diffusion Model for High-quality and Realistic 3D Human Generation. 4568-4577
- Hongchi Xia, Zhi-Hao Lin, Wei-Chiu Ma, Shenlong Wang:
Video2Game: Real-time, Interactive, Realistic and Browser-Compatible Environment from a Single Video. 4578-4588
- Vikas Thamizharasan, Difan Liu, Matthew Fisher, Nanxuan Zhao, Evangelos Kalogerakis, Michal Lukáč:
NIVeL: Neural Implicit Vector Layers for Text-to-Vector Generation. 4589-4597
- Jinseo Jeong, Junseo Koo, Qimeng Zhang, Gunhee Kim:
ESR-NeRF: Emissive Source Reconstruction Using LDR Multi-View Images. 4598-4609
- Linqi Zhou, Andy Shih, Chenlin Meng, Stefano Ermon:
DreamPropeller: Supercharge Text-to-3D Generation with Parallel Sampling. 4610-4619
- Chenjian Gao, Boyan Jiang, Xinghui Li, Yingpeng Zhang, Qian Yu:
GenesisTex: Adapting Image Denoising Diffusion to Texture Space. 4620-4629



- Lior Yariv, Omri Puny, Oran Gafni, Yaron Lipman:
Mosaic-SDF for 3D Generative Models. 4630-4639
- Michael Fischer, Zhengqin Li, Thu Nguyen-Phuoc, Aljaz Bozic, Zhao Dong, Carl S. Marshall, Tobias Ritschel:
NeRF Analogies: Example-Based Visual Attribute Transfer for NeRFs. 4640-4650
- Xingtao Wang, Hongliang Wei, Xiaopeng Fan, Debin Zhao:
Hyper-MD: Mesh Denoising with Customized Parameters Aware of Noise Intensity and Geometric Characteristics. 4651-4660
- Maximilian Frühauf, Hayko Riemenschneider, Markus Gross, Christopher Schroers:
QUADify: Extracting Meshes with Pixel-Level Details and Materials from Images. 4661-4670
- Pu Li, Jianwei Guo, Huibin Li, Bedrich Benes, Dong-Ming Yan:
SfmCAD: Unsupervised CAD Reconstruction by Learning Sketch-based Feature Modeling Operations. 4671-4680
- Ramana Sundararaman, Roman Klokov, Maks Ovsjanikov:
Self-Supervised Dual Contouring. 4681-4691
- Yuan Li, Zhihao Liu, Bedrich Benes, Xiaopeng Zhang, Jianwei Guo:
SVTree: Semantic Voxel Diffusion for Single Image Tree Reconstruction. 4692-4702
- Vanessa Sklyarova, Egor Zakharov, Otmar Hilliges, Michael J. Black, Justus Thies:
Text-Conditioned Generative Model of 3D Strand-Based Human Hairstyles. 4703-4712
- Mohammad Sadil Khan, Elona Dupont, Sk Aziz Ali, Kseniya Cherenkova, Anis Kacem, Djamila Aouada:
CAD-SIGNet: CAD Language Inference from Point Clouds Using Layer-Wise Sketch Instance Guided Attention. 4713-4722
- Biao Zhang, Peter Wonka:
Functional Diffusion. 4723-4732
- Chenyang Si, Ziqi Huang, Yuming Jiang, Ziwei Liu:
FreeU: Free Lunch in Diffusion U-Net. 4733-4743
- Yutong Feng, Biao Gong, Di Chen, Yujun Shen, Yu Liu, Jingren Zhou:
Ranni: Taming Text-to-Image Diffusion for Accurate Instruction Following. 4744-4753
- Hexiang Hu, Kelvin C. K. Chan, Yu-Chuan Su, Wenhui Chen, Yandong Li, Kihyuk Sohn, Yang Zhao, Xue Ben, Boqing Gong, William W. Cohen, Ming-Wei Chang, Xuhui Jia:
Instruct-Imagen: Image Generation with Multi-modal Instruction. 4754-4763
- Yanbing Zhang, Mengping Yang, Qin Zhou, Zhe Wang:
Attention Calibration for Disentangled Text-to-Image Personalization. 4764-4774
- Amir Hertz, Andrey Voynov, Shlomi Fruchter, Daniel Cohen-Or:
Style Aligned Image Generation via Shared Attention. 4775-4785
- Damien Teney, Armand Mihai Nicolicioiu, Valentin Hartmann, Ehsan Abbasnejad:
Neural Redshift: Random Networks are not Random Functions. 4786-4796
- Runpeng Yu, Xinchao Wang:
Neural Lineage. 4797-4807
- Lucas Brynte, José Pedro Iglesias, Carl Olsson, Fredrik Kahl:
Learning Structure-From-Motion with Graph Attention Networks. 4808-4817
- Bin Xiao, Haiping Wu, Weijian Xu, Xiyang Dai, Houdong Hu, Yumao Lu, Michael Zeng, Ce Liu, Lu Yuan:
Florence-2: Advancing a Unified Representation for a Variety of Vision Tasks. 4818-4829
- Junwon Seo, Sangyoon Lee, Kwang In Kim, Jaeho Lee:
In Search of a Data Transformation that Accelerates Neural Field Training. 4830-4839
- Xiaoyang Wu, Li Jiang, Peng-Shuai Wang, Zhijian Liu, Xihui Liu, Yu Qiao, Wanli Ouyang, Tong He, Hengshuang Zhao:
Point Transformer V3: Simpler, Faster, Stronger. 4840-4851
- Axel Barroso-Laguna, Sowmya Munukutla, Victor Adrian Prisacariu, Eric Brachmann:
Matching 2D Images in 3D: Metric Relative Pose from Metric Correspondences. 4852-4863
- Hadi Alzayer, Kevin Zhang, Brandon Y. Feng, Christopher A. Metzler, Jia-Bin Huang:
Seeing the World through Your Eyes. 4864-4873
- Zhiqiang Yan, Yuankai Lin, Kun Wang, Yupeng Zheng, Yufei Wang, Zhenyu Zhang, Jun Li, Jian Yang:
Tri-Perspective view Decomposition for Geometry-Aware Depth Completion. 4874-4884
- Georg Bökman, Johan Edstedt, Michael Felsberg, Fredrik Kahl:
Steerers: A Framework for Rotation Equivariant Keypoint Descriptors. 4885-4895
- Yang Chen, Yingwei Pan, Haibo Yang, Ting Yao, Tao Mei:
VP3D: Unleashing 2D Visual Prompt for Text-to-3D Generation. 4896-4905
- Zhiyuan Min, Yawei Luo, Wei Yang, Yuesong Wang, Yi Yang:
Entangled View-Epipolar Information Aggregation for Generalizable Neural Radiance Fields. 4906-4916
- Chengyao Wang, Li Jiang, Xiaoyang Wu, Zhuotao Tian, Bohao Peng, Hengshuang Zhao, Jiaya Jia:
GroupContrast: Semantic-Aware Self-Supervised Representation Learning for 3D Understanding. 4917-4928



- Yu Meng, Zhou Xue, Xu Chang, Xuemei Hu, Tao Yue:
iToF-Flow-Based High Frame Rate Depth Imaging. 4929-4938
- Haechan Lee, Wonjoon Jin, Seung-Hwan Baek, Sunghyun Cho:
Generalizable Novel-View Synthesis Using a Stereo Camera. 4939-4948
- Zhipeng Hu, Minda Zhao, Chaoyi Zhao, Xinyue Liang, Lincheng Li, Zeng Zhao, Changjie Fan, Xiaowei Zhou, Xin Yu:
EfficientDreamer: High-Fidelity and Stable 3D Creation via Orthogonal-view Diffusion Priors. 4949-4958
- Lalit Manam, Venu Madhav Govindu:
Leveraging Camera Triplets for Efficient and Accurate Structure-from-Motion. 4959-4968
- Lukas Radl, Michael Steiner, Andreas Kurz, Markus Steinberger:
LAENeRF: Local Appearance Editing for Neural Radiance Fields. 4969-4978
- Kirill Mazur, Gwangbin Bae, Andrew J. Davison:
SuperPrimitive: Scene Reconstruction at a Primitive Level. 4979-4989
- Felix Rydell, Angélica Torres, Viktor Larsson:
Revisiting Sampson Approximations for Geometric Estimation Problems. 4990-4998
- Shaocong Dong, Lihe Ding, Zhanpeng Huang, Zibin Wang, Tianfan Xue, Dan Xu:
Interactive3D: Create What You Want by Interactive 3D Generation. 4999-5008
- Zihan Gao, Licheng Jiao, Lingling Li, Xu Liu, Fang Liu, Puhua Chen, Yuwei Guo:
Multiplane Prior Guided Few-Shot Aerial Scene Rendering. 5009-5019
- Zhiyin Qian, Shaofei Wang, Marko Mihajlovic, Andreas Geiger, Siyu Tang:
3DGSAvatar: Animatable Avatars via Deformable 3D Gaussian Splatting. 5020-5030
- Ange Lou, Benjamin Planche, Zhongpai Gao, Yamin Li, Tianyu Luan, Hao Ding, Terrence Chen, Jack H. Noble, Ziyan Wu:
DaReNeRF: Direction-aware Representation for Dynamic Scenes. 5031-5042
- Lukas Höllerin, Aljaz Bozic, Norman Müller, David Novotný, Hung-Yu Tseng, Christian Richardt, Michael Zollhöfer, Matthias Nießner:
ViewDiff: 3D-Consistent Image Generation with Text-to-Image Models. 5043-5052
- Jaehoon Choi, Rajvi Shah, Qinbo Li, Yipeng Wang, Ayush Saraf, Changil Kim, Jia-Bin Huang, Dinesh Manocha, Suhib Alsian, Johannes Kopf:
LTM: Lightweight Textured Mesh Extraction and Refinement of Large Unbounded Scenes for Efficient Storage and Real-Time Rendering. 5053-5063
- Andrea Porfiri Dal Cin, Timothy Duff, Luca Magri, Tomás Pajdla:
Minimal Perspective Autocalibration. 5064-5073
- Shuofeng Sun, Yongming Rao, Jiwen Lu, Haibin Yan:
X-3D: Explicit 3D Structure Modeling for Point Cloud Recognition. 5074-5083
- Junkai Deng, Fei Hou, Xuhui Chen, Wencheng Wang, Ying He:
2S-UDF: A Novel Two-Stage UDF Learning Method for Robust Non-Watertight Model Reconstruction from Multi-View Images. 5084-5093
- Youngju Na, Woo Jae Kim, Kyu Beom Han, Suhyeon Ha, Sung-Eui Yoon:
UFORecon: Generalizable Sparse-View Surface Reconstruction from Arbitrary and Unfavorable Sets. 5094-5104
- Xiangyue Liu, Han Xue, Kunming Luo, Ping Tan, Li Yi:
GenN2N: Generative NeRF2NeRF Translation. 5105-5114
- Lihe Ding, Shaocong Dong, Zhanpeng Huang, Zibin Wang, Yiyuan Zhang, Kaixiong Gong, Dan Xu, Tianfan Xue:
Text-to-3D Generation with Bidirectional Diffusion Using Both 2D and 3D Priors. 5115-5124
- Yaqing Ding, Jonathan Astermark, Magnus Oskarsson, Viktor Larsson:
Noisy One-Point Homographies are Surprisingly Good. 5125-5134
- Peng Xu, Zhiyu Xiang, Chengyu Qiao, Jingyun Fu, Tianyu Pu:
Adaptive Multi-Modal Cross-Entropy Loss for Stereo Matching. 5135-5144
- Zehan Zheng, Fan Lu, Weiyi Xue, Guang Chen, Changjun Jiang:
LiDAR4D: Dynamic Neural Fields for Novel Space-Time View LiDAR Synthesis. 5145-5154
- Ziyi Chen, Xiaolong Wu, Yu Zhang:
NC-SDF: Enhancing Indoor Scene Reconstruction Using Neural SDFs with View-Dependent Normal Compensation. 5155-5165
- Jiaqi Lin, Zhihao Li, Xiao Tang, Jianzhuang Liu, Shiyong Liu, Jiayue Liu, Yangdi Lu, Xiaofei Wu, Songcen Xu, Youliang Yan, Wenming Yang:
VastGaussian: Vast 3D Gaussians for Large Scene Reconstruction. 5166-5175
- Ka-Chun Shum, Jaeyeon Kim, Bin-Son Hua, Duc Thanh Nguyen, Sai-Kit Yeung:
Language-driven Object Fusion into Neural Radiance Fields with Pose-Conditioned Dataset Updates. 5176-5187
- Yanzhe Liu, Rong Chen, Yushi Li, Yixi Li, Xuehou Tan:
SPU-PMD: Self-Supervised Point Cloud Upsampling via Progressive Mesh Deformation. 5188-5197
- Peter Kocsis, Vincent Sitzmann, Matthias Nießner:
Intrinsic Image Diffusion for Indoor Single-view Material Estimation. 5198-5208



- Zicheng Zhang, Ruobing Zheng, Bonan Li, Congying Han, Tianqi Li, Meng Wang, Tiande Guo, Jingdong Chen, Ziwen Liu, Ming Yang:
Learning Dynamic Tetrahedra for High-Quality Talking Head Synthesis. 5209-5219
- Viktor Kocur, Daniel Kyselica, Zuzana Kukelova:
Robust Self-Calibration of Focal Lengths from the Fundamental Matrix. 5220-5229
- Baptiste Brument, Robin Bruneau, Yvain Quéau, Jean Mélou, François Bernard Lauze, Jean-Denis Durou, Lilian Calvet:
RNb-NeuS: Reflectance and Normal-Based Multi-View 3D Reconstruction. 5230-5239
- Hao-Bin Duan, Miao Wang, Yan-Xun Li, Yong-Liang Yang:
Neural 3D Strokes: Creating Stylized 3D Scenes with Vectorized 3D Strokes. 5240-5249
- Jiacheng Deng, Jiahao Lu, Tianzhu Zhang:
Unsupervised Template-assisted Point Cloud Shape Correspondence Network. 5250-5259
- Shaohan Li, Yunpeng Shi, Gilad Lerman:
Efficient Detection of Long Consistent Cycles and its Application to Distributed Synchronization. 5260-5269
- Jamie Watson, Filippo Aleotti, Mohamed Sayed, Zawar Qureshi, Oisin Mac Aodha, Gabriel J. Brostow, Michael Firman, Sara Vicente:
AirPlanes: Accurate Plane Estimation via 3D-Consistent Embeddings. 5270-5280
- Jonas Kälble, Sascha Wirges, Maxim Tatarchenko, Eddy Ilg:
Accurate Training Data for Occupancy Map Prediction in Automated Driving Using Evidence Theory. 5281-5290
- Qi Ma, Danda Pani Paudel, Ajad Chhatkuli, Luc Van Gool:
Continuous Pose for Monocular Cameras in Neural Implicit Representation. 5291-5301
- Fangzhou Mu, Carter Sifferman, Sacha Jungerman, Yiquan Li, Mark Han, Michael Gleicher, Mohit Gupta, Yin Li:
Towards 3D Vision with Low-Cost Single-Photon Cameras. 5302-5311
- Yongzhe Yuan, Yue Wu, Xiaolong Fan, Maoguo Gong, Qiguang Miao, Wenping Ma:
Inlier Confidence Calibration for Point Cloud Registration. 5312-5321
- Yingwenqi Jiang, Jiadong Tu, Yuan Liu, Xifeng Gao, Xiaoxiao Long, Wenping Wang, Yuexin Ma:
GaussianShader: 3D Gaussian Splatting with Shading Functions for Reflective Surfaces. 5322-5332
- Jin-Chuan Shi, Miao Wang, Hao-Bin Duan, Shao-Hua Guan:
Language Embedded 3D Gaussians for Open-Vocabulary Scene Understanding. 5333-5343
- Honghua Chen, Chen Change Loy, Xingang Pan:
MViP-NeRF: Multi-View 3D Inpainting on NeRF Scenes via Diffusion Prior. 5344-5353
- Antoine Guédon, Vincent Lepetit:
SuGaR: Surface-Aligned Gaussian Splatting for Efficient 3D Mesh Reconstruction and High-Quality Mesh Rendering. 5354-5363
- Tianyu Huang, Yihan Zeng, Zhilu Zhang, Wan Xu, Hang Xu, Songcen Xu, Rynson W. H. Lau, Wangmeng Zuo:
DreamControl: Control-Based Text-to-3D Generation with 3D Self-Prior. 5364-5373
- Silvia Zuffi, Ylva Mellbin, Ci Li, Markus Höschle, Hedvig Kjellström, Senya Polikovsky, Elin Hernlund, Michael J. Black:
VAREN: Very Accurate and Realistic Equine Network. 5374-5383
- Chaoyue Song, Jiacheng Wei, Chuan Sheng Foo, Guosheng Lin, Fayao Liu:
REACTO: Reconstructing Articulated Objects from a Single Video. 5384-5395
- Jaehyeok Shim, Kyungdon Joo:
DITTO: Dual and Integrated Latent Topologies for Implicit 3D Reconstruction. 5396-5405
- Weiyao Wang, Pierre Gleize, Hao Tang, Xingyu Chen, Kevin J. Liang, Matt Feiszli:
ICON: Incremental CONfidence for Joint Pose and Radiance Field Optimization. 5406-5417
- Yiyang Chen, Lunhao Duan, Shanshan Zhao, Changxing Ding, Dacheng Tao:
Local-consistent Transformation Learning for Rotation-invariant Point Cloud Analysis. 5418-5427
- Xiao Tang, Min Yang, Penghui Sun, Hui Li, Yuchao Dai, Feng Zhu, Hojae Lee:
PaReNeRF: Toward Fast Large-Scale Dynamic NeRF with Patch-Based Reference. 5428-5438
- Gabriel Dogadov, Ugo Paavo Finnendahl, Marc Alexa:
Fitting Flats to Flats. 5439-5447
- Marco Pesavento, Yuanlu Xu, Nikolaos Sarafianos, Robert Maier, Ziyan Wang, Chun-Han Yao, Marco Volino, Edmond Boyer, Adrian Hilton, Tony Tung:
ANIM: Accurate Neural Implicit Model for Human Reconstruction from a Single RGB-D Image. 5448-5458
- Tongfan Guan, Chen Wang, Yun-Hui Liu:
Neural Markov Random Field for Stereo Matching. 5459-5469
- Takuhiro Kaneko:
Improving Physics-Augmented Continuum Neural Radiance Field-Based Geometry-Agnostic System Identification with Lagrangian Particle Optimization. 5470-5480
- Tobias Kirschstein, Simon Giebenhain, Matthias Nießner:
DiffusionAvatars: Deferred Diffusion for High-fidelity 3D Head Avatars. 5481-5492
- Chunlong Xia, Xinliang Wang, Feng Lv, Xin Hao, Yifeng Shi:
ViT-CoMer: Vision Transformer with Convolutional Multi-scale Feature Interaction for Dense Predictions. 5493-



5502

- Ruixuan Yu, Jian Sun:
Pose-Transformed Equivariant Network for 3D Point Trajectory Prediction. 5503-5512
- Xiaohan Ding, Yiyuan Zhang, Yixiao Ge, Sijie Zhao, Lin Song, Xiangyu Yue, Ying Shan:
UniRepLKNet: A Universal Perception Large-Kernel ConvNet for Audio, Video, Point Cloud, Time-Series and Image Recognition. 5513-5524
- Hugues Thomas, Yao-Hung Hubert Tsai, Timothy D. Barfoot, Jian Zhang:
KPConvX: Modernizing Kernel Point Convolution with Kernel Attention. 5525-5535
- Otniel-Bogdan Mercea, Alexey A. Gritsenko, Cordelia Schmid, Anurag Arnab:
Time-, Memory- and Parameter-Efficient Visual Adaptation. 5536-5545
- Yikang Li, Yeqing Qiu, Yuxuan Chen, Lingshen He, Zhouchen Lin:
Affine Equivariant Networks Based on Differential Invariants. 5546-5556
- Honghao Chen, Xiangxiang Chu, Yongjian Ren, Xin Zhao, Kaiqi Huang:
PeLK: Parameter-Efficient Large Kernel ConvNets with Peripheral Convolution. 5557-5567
- Renan A. Rojas-Gomez, Teck-Yian Lim, Minh N. Do, Raymond A. Yeh:
Making Vision Transformers Truly Shift-Equivariant. 5568-5577
- Hancheng Ye, Chong Yu, Peng Ye, Renqiu Xia, Yansong Tang, Jiwen Lu, Tao Chen, Bo Zhang:
Once for Both: Single Stage of Importance and Sparsity Search for Vision Transformer Compression. 5578-5588
- Chunxiao Fan, Ziqi Wang, Dan Guo, Meng Wang:
Data-Free Quantization via Pseudo-label Filtering. 5589-5598
- Yuxiang Lu, Suizhi Huang, Yuwen Yang, Shalayiding Sirejiding, Yue Ding, Hongtao Lu:
Fedhca²: Towards Hetero-Client Federated Multi-Task Learning. 5599-5609
- Xinyu Shi, Zecheng Hao, Zhaofei Yu:
SpikingResformer: Bridging ResNet and Vision Transformer in Spiking Neural Networks. 5610-5619
- Lewei Yao, Renjie Pi, Jianhua Han, Xiaodan Liang, Hang Xu, Wei Zhang, Zhenguo Li, Dan Xu:
DetCLIPv3: Towards Versatile Generative Open-Vocabulary Object Detection. 5610-5619
- Pavlo Melnyk, Andreas Robinson, Michael Felsberg, Mårten Wadenbäck:
TetraSphere: A Neural Descriptor for O(3)-Invariant Point Cloud Analysis. 5620-5630
- Tao Li, Pan Zhou, Zhengbao He, Xinwen Cheng, Xiaolin Huang:
Friendly Sharpness-Aware Minimization. 5631-5640
- Qihang Fan, Huabo Huang, Mingrui Chen, Hongmin Liu, Ran He:
RMT: Retentive Networks Meet Vision Transformers. 5641-5651
- Yuwen Xiong, Zhiqi Li, Yuntao Chen, Feng Wang, Xizhou Zhu, Jiapeng Luo, Wenhai Wang, Tong Lu, Hongsheng Li, Yu Qiao, Lewei Lu, Jie Zhou, Jifeng Dai:
Efficient Deformable ConvNets: Rethinking Dynamic and Sparse Operator for Vision Applications. 5652-5661
- Beichen Zhang, Xiaoxing Wang, Xiaohan Qin, Junchi Yan:
Boosting Order-Preserving and Transferability for Neural Architecture Search: A Joint Architecture Refined Search and Fine-Tuning Approach. 5662-5671
- Weihao Yu, Pan Zhou, Shuicheng Yan, Xinchao Wang:
InceptionNeXt: When Inception Meets ConvNeXt. 5672-5683
- Edwin Vargas, Claudia V. Correa P., Carlos Hinojosa, Henry Arguello:
BiPer: Binary Neural Networks Using a Periodic Function. 5684-5693
- Xu Ma, Xiyang Dai, Yue Bai, Yizhou Wang, Yun Fu:
Rewrite the Stars. 5694-5703
- Ruichen Ma, Guanchao Qiao, Yian Liu, Liwei Meng, Ning Ning, Yang Liu, Shaogang Hu:
A&B BNN: Add&Bit-Operation-Only Hardware-Friendly Binary Neural Network. 5704-5713
- Guikun Chen, Xia Li, Yi Yang, Wenguan Wang:
Neural Clustering Based Visual Representation Learning. 5714-5725
- Keith G. Mills, Fred X. Han, Mohammad Salameh, Shengyao Lu, Chunhua Zhou, Jiao He, Fengyu Sun, Di Niu:
Building Optimal Neural Architectures Using Interpretable Knowledge. 5726-5735
- Mengfei Xia, Yujun Shen, Changsong Lei, Yu Zhou, Deli Zhao, Ran Yi, Wenping Wang, Yong-Jin Liu:
Towards More Accurate Diffusion Model Acceleration with a Timestep Tuner. 5736-5745
- Jingjing Xie, Yuxin Zhang, Mingbao Lin, Zhihang Lin, Liujuan Cao, Rongrong Ji:
UniPTS: A Unified Framework for Proficient Post-Training Sparsity. 5746-5755
- Seokju Yun, Youngmin Ro:
SHViT: Single-Head Vision Transformer with Memory Efficient Macro Design. 5756-5767
- Aihua Mao, Biao Yan, Zijing Ma, Ying He:
Denoising Point Clouds in Latent Space via Graph Convolution and Invertible Neural Network. 5768-5777
- Weiying Xie, Haowei Li, Jitao Ma, Yunsong Li, Jie Lei, Donglai Liu, Leyuan Fang:
JointSQ: Joint Sparsification-Quantization for Distributed Learning. 5778-5787



- Alon Zolfi, Guy Amit, Amit Baras, Satoru Koda, Ikuya Morikawa, Yuval Elovici, Asaf Shabtai:
YOLOOD: Utilizing Object Detection Concepts for Multi-Label Out-of-Distribution Detection. 5788-5797
- Xiang Fei, Xiawu Zheng, Yan Wang, Fei Chao, Chenglin Wu, Liujuan Cao:
RepAn: Enhanced Annealing through Re-parameterization. 5798-5808
- Duo Su, Junjie Hou, Weizhi Gao, Yingjie Tian, Bowen Tang:
D⁴M: Dataset Distillation via Disentangled Diffusion Model. 5809-5818
- Fangjinhua Wang, Xudong Jiang, Silvano Galliani, Christoph Vogel, Marc Pollefeys:
GLACE: Global Local Accelerated Coordinate Encoding. 5819-5828
- Nikola Zubic, Mathias Gehrig, Davide Scaramuzza:
State Space Models for Event Cameras. 5819-5828
- Sofia Casarin, Cynthia Ifeyinwa Ugwu, Sergio Escalera, Oswald Lanz:
Your Image Is My Video: Reshaping the Receptive Field via Image-to-Video Differentiable AutoAugmentation and Fusion. 5829-5839
- Tahira Shehzadi, Khurram Azeem Hashmi, Didier Stricker, Muhammad Zeshan Afzal:
Sparse Semi-DETR: Sparse Learnable Queries for Semi-Supervised Object Detection. 5840-5850
- Xuzhe Zhang, Yuhao Wu, Elsa D. Angelini, Ang Li, Jia Guo, Jerod M. Rasmussen, Thomas G. O'Connor, Pathik D. Wadhwa, Andrea Parolin Jackowski, Hai Li, Jonathan Posner, Andrew F. Laine, Yun Wang:
MAPSeg: Unified Unsupervised Domain Adaptation for Heterogeneous Medical Image Segmentation Based on 3D Masked Autoencoding and Pseudo-Labeling. 5851-5862
- Ha Min Son, Moon-Hyun Kim, Tai-Myoung Chung, Chao Huang, Xin Liu:
FedUV: Uniformity and Variance for Heterogeneous Federated Learning. 5863-5872
- Ashish Kumar, Daneul Kim, Jaesik Park, Laxmidhar Behera:
Pick-or-Mix: Dynamic Channel Sampling for ConvNets. 5873-5882
- Zhiyuan Yu, Li Shen, Liang Ding, Xinmei Tian, Yixin Chen, Dacheng Tao:
Sheared Backpropagation for Fine-Tuning Foundation Models. 5883-5892
- Junghyup Lee, Bumsub Ham:
AZ-NAS: Assembling Zero-Cost Proxies for Network Architecture Search. 5893-5903
- Sumanth Udupa, Prajwal Gurunath, Aniruddh Sikdar, Suresh Sundaram:
MRFP: Learning Generalizable Semantic Segmentation from Sim-2-Real with Multi-Resolution Feature Perturbation. 5904-5914
- Zhengqi Xu, Ke Yuan, Huiqiong Wang, Yong Wang, Mingli Song, Jie Song:
Training-Free Pretrained Model Merging. 5915-5925
- Cansu Korkmaz, A. Murat Tekalp, Zafer Dogan:
Training Generative Image Super-Resolution Models by Wavelet-Domain Losses Enables Better Control of Artifacts. 5926-5936
- Alessio Mazzucchelli, Adrian Garcia-Garcia, Elena Garces, Fernando Rivas-Manzaneque, Francesc Moreno-Noguer, Adrián Peñate Sánchez:
IReNe: Instant Recoloring of Neural Radiance Fields. 5937-5946
- Sudong Cai:
AdaShift: Learning Discriminative Self-Gated Neural Feature Activation With an Adaptive Shift Factor. 5947-5956
- Jinzhi Zheng, Heng Fan , Libo Zhang:
Kernel Adaptive Convolution for Scene Text Detection via Distance Map Prediction. 5957-5966
- Yuwei Ou, Yuqi Feng, Yanan Sun:
Towards Accurate and Robust Architectures via Neural Architecture Search. 5967-5976
- Jinfeng Xu, Siyuan Yang, Xianzhi Li, Yuan Tang, Yixue Hao, Long Hu, Min Chen:
PDF: A Probability-Driven Framework for Open World 3D Point Cloud Semantic Segmentation. 5977-5986
- Hengyuan Xu, Liyao Xiang, Hangyu Ye, Duxi Yao, Pengzhi Chu, Baochun Li:
Permutation Equivariance of Transformers and its Applications. 5987-5996
- Hyejin Park, Jeongyeon Hwang, Sunung Mun, Sangdon Park, Jungseul Ok:
MedBN: Robust Test-Time Adaptation against Malicious Test Samples. 5997-6007
- He Liu, Yikai Wang, Huaping Liu, Fuchun Sun, Anbang Yao:
Small Scale Data-Free Knowledge Distillation. 6008-6016
- Kosuke Sumiyasu, Kazuhiko Kawamoto, Hiroshi Kera:
Identifying Important Group of Pixels using Interactions. 6017-6026
- Khiem Le, Long Ho, Cuong Do, Danh Le Phuoc, Kok-Seng Wong:
Efficiently Assemble Normalization Layers and Regularization for Federated Domain Generalization. 6027-6036
- Xinyu Geng, Jiaming Wang, Jiawei Gong, Yuerong Xue, Jun Xu, Fanglin Chen, Xiaolin Huang:
OrthCaps: An Orthogonal CapsNet with Sparse Attention Routing and Pruning. 6037-6046
- Takumi Kobayashi:
Mean-Shift Feature Transformer. 6047-6056



- Shuxi Zhang, Hanpeng Liu, Stephen Lin, Kun He:
You Only Need Less Attention at Each Stage in Vision Transformers. 6057-6066
- Oscar Carlsson, Jan E. Gerken, Hampus Linander, Heiner Spieß, Fredrik Ohlsson, Christoffer Petersson, Daniel Persson:
HEAL-SWIN: A Vision Transformer on the Sphere. 6067-6077
- David Osowiechi, Gustavo Adolfo Vargas Hakim, Mehrdad Noori, Milad Cheraghalkhani, Ali Bahri, Moslem Yazdanpanah, Ismail Ben Ayed, Christian Desrosiers:
NC-TTT: A Noise Contrastive Approach for Test-Time Training. 6078-6086
- Wenlong Deng, Christos Thrampoulidis, Xiaoxiao Li:
Unlocking the Potential of Prompt-Tuning in Bridging Generalized and Personalized Federated Learning. 6087-6097
- Siddharth Roheda, Amit Satish Unde, Loay Rashid:
MR-VNet: Media Restoration using Volterra Networks. 6098-6107
- Sherry X. Chen, Yaron Vaxman, Elad Ben Baruch, David Asulin, Aviad Moreshet, Kuo-Chin Lien, Misha Sra, Pradeep Sen:
TiNO-Edit: Timestep and Noise Optimization for Robust Diffusion-Based Image Editing. 6108-6117
- Yiyuan Zhang, Xiaohan Ding, Kaixiong Gong, Yixiao Ge, Ying Shan, Xiangyu Yue:
Multimodal Pathway: Improve Transformers with Irrelevant Data from Other Modalities. 6108-6117
- Mustafa Munir, William Avery, Md Mostafijur Rahman, Radu Marculescu:
GreedyViG: Dynamic Axial Graph Construction for Efficient Vision GNNs. 6118-6127
- Dongyeong Hwang, Hyunju Kim, Sunwoo Kim, Kijung Shin:
FlowerFormer: Empowering Neural Architecture Encoding Using a Flow-Aware Graph Transformer. 6128-6137
- Huancheng Chen, Haris Vikalo:
Mixed-Precision Quantization for Federated Learning on Resource-Constrained Heterogeneous Devices. 6138-6148
- Zhiyu Qu, Lan Yang, Honggang Zhang, Tao Xiang, Kaiyue Pang, Yi-Zhe Song:
Wired Perspectives: Multi-View Wire Art Embraces Generative AI. 6149-6158
- Ruoyi Du, Dongliang Chang, Timothy M. Hospedales, Yi-Zhe Song, Zhanyu Ma:
DemoFusion: Democratising High-Resolution Image Generation With No \$\$\$. 6159-6168
- Chenyang Wang, Zerong Zheng, Tao Yu, Xiaoqian Lv, Bineng Zhong, Shengping Zhang, Liqiang Nie:
DiffPerformer: Iterative Learning of Consistent Latent Guidance for Diffusion-Based Human Video Generation. 6169-6179
- Jiun Tian Hoe, Xudong Jiang, Chee Seng Chan, Yap-Peng Tan, Weipeng Hu:
InteractDiffusion: Interaction Control in Text-to-Image Diffusion Models. 6180-6189
- Chang Liu, Haoning Wu, Yujie Zhong, Xiaoyun Zhang, Yanfeng Wang, Weidi Xie:
Intelligent Grimm - Open-ended Visual Storytelling via Latent Diffusion Models. 6190-6200
- Jonas Schult, Sam S. Tsai, Lukas Höller, Bichen Wu, Jialiang Wang, Chih-Yao Ma, Kunpeng Li, Xiaofang Wang, Felix Wimbauer, Zijian He, Peizhao Zhang, Bastian Leibe, Peter Vajda, Ji Hou:
ControlRoom3D: Room Generation Using Semantic Proxy Rooms. 6201-6210
- Felix Wimbauer, Bichen Wu, Edgar Schönfeld, Xiaoliang Dai, Ji Hou, Zijian He, Arsiom Sanakoyeu, Peizhao Zhang, Sam S. Tsai, Jonas Kohler, Christian Rupprecht, Daniel Cremers, Peter Vajda, Jialiang Wang:
Cache Me if You Can: Accelerating Diffusion Models through Block Caching. 6211-6220
- Ziqi Cai, Kaiwen Jiang, Shu-Yu Chen, Yu-Kun Lai, Hongbo Fu, Boxin Shi, Lin Gao:
Real-Time 3D-Aware Portrait Video Relighting. 6221-6231
- Xudong Wang, Trevor Darrell, Sai Saketh Rambhatla, Rohit Girdhar, Ishan Misra:
InstanceDiffusion: Instance-Level Control for Image Generation. 6232-6242
- Junshu Tang, Yanhong Zeng, Ke Fan, Xuheng Wang, Bo Dai, Kai Chen, Lizhuang Ma:
Make-It-Vivid: Dressing Your Animatable Biped Cartoon Characters from Text. 6243-6253
- Shanglin Li, Bohan Zeng, Yutang Feng, Sicheng Gao, Xiuhui Liu, Jiaming Liu, Lin Li, Xu Tang, Yao Hu, Jianzhuang Liu, Baochang Zhang:
ZONE: Zero-Shot Instruction-Guided Local Editing. 6254-6263
- Nicolas Dufour, Victor Besnier, Vicky Kalogeiton, David Picard:
Don't Drop Your Samples! Coherence-Aware Training Benefits Conditional Diffusion. 6264-6273
- Sachit Menon, Ishan Misra, Rohit Girdhar:
Generating Illustrated Instructions. 6274-6284
- Lin Zhu, Kangmin Jia, Yifan Zhao, Yunshan Qi, Lizhi Wang, Hua Huang:
SpikeNeRF: Learning Neural Radiance Fields from Continuous Spike Stream. 6285-6295
- Ziyu Wang, Yue Xu, Cewu Lu, Yong-Lu Li:
Dancing with Still Images: Video Distillation via Static-Dynamic Disentanglement. 6296-6304
- Lu Qi, Lehan Yang, Weidong Guo, Yu Xu, Bo Du, Varun Jampani, Ming-Hsuan Yang:
UniGS: Unified Representation for Image Generation and Segmentation. 6305-6315
- Kilichbek Haydarov, Aashiq Muhamed, Xiaoqian Shen, Jovana Lazarevic, Ivan Skorokhodov, Chamuditha Jayanga Galappaththige, Mohamed Elhoseiny:
Adversarial Text to Continuous Image Generation. 6316-6326



- Tsung-Han Wu, Long Lian, Joseph E. Gonzalez, Boyi Li, Trevor Darrell:
Self-Correcting LLM-Controlled Diffusion Models. 6327-6336
- Cheng Zhang, Qianyi Wu, Camilo Cruz Gambardella, Xiaoshui Huang, Dinh Phung, Wanli Ouyang, Jianfei Cai:
Taming Stable Diffusion for Text to 360° Panorama Image Generation. 6347-6357
- Jingyuan Yang, Jiawei Feng, Hui Huang:
EmoGen: Emotional Image Content Generation with Text-to-Image Diffusion Models. 6358-6368
- Desai Xie, Jiahao Li, Hao Tan, Xin Sun, Zhixin Shu, Yi Zhou, Sai Bi, Sören Pirk, Arie E. Kaufman:
Carve3D: Improving Multi-view Reconstruction Consistency for Diffusion Models with RL Finetuning. 6369-6379
- Jiawei Ren, Mengmeng Xu, Jui-Chieh Wu, Ziwei Liu, Tao Xiang, Antoine Toisoul:
Move Anything with Layered Scene Diffusion. 6380-6389
- Lirui Zhao, Yue Yang, Kaipeng Zhang, Wenqi Shao, Yuxin Zhang, Yu Qiao, Ping Luo, Rongrong Ji:
DiffAgent: Fast and Accurate Text-to-Image API Selection with Large Language Model. 6390-6399
- Chao Liang, Fan Ma, Linchao Zhu, Yingying Deng, Yi Yang:
CapHuman: Capture Your Moments in Parallel Universes. 6400-6409
- Mengshun Hu, Kui Jiang, Zhihang Zhong, Zheng Wang, Yinqiang Zheng:
IQ-VFI: Implicit Quadratic Motion Estimation for Video Frame Interpolation. 6410-6419
- Yanzuo Lu, Manlin Zhang, Andy J. Ma, Xiaohua Xie, Jianhuang Lai:
Coarse-to-Fine Latent Diffusion for Pose-Guided Person Image Synthesis. 6420-6429
- Shilin Lu, Zilan Wang, Leyang Li, Yanzhu Liu, Adams Wai-Kin Kong:
MACE: Mass Concept Erasure in Diffusion Models. 6430-6440
- Shoufa Chen, Mengmeng Xu, Jiawei Ren, Yuren Cong, Sen He, Yanping Xie, Animesh Sinha, Ping Luo, Tao Xiang, Juan-Manuel Pérez-Rúa:
GenTron: Diffusion Transformers for Image and Video Generation. 6441-6451
- Mengwei Ren, Wei Xiong, Jae Shin Yoon, Zhixin Shu, Jianming Zhang, Hyunjoo Jung, Guido Gerig, He Zhang:
Relightful Harmonization: Lighting-Aware Portrait Background Replacement. 6452-6462
- Hangjie Yuan, Shiwei Zhang, Xiang Wang, Yujie Wei, Tao Feng, Yining Pan, Yingya Zhang, Ziwei Liu, Samuel Albanie, Dong Ni:
InstructVideo: Instructing Video Diffusion Models with Human Feedback. 6463-6474
- Minye Wu, Zehao Wang, Georgios Kouros, Tinne Tuytelaars:
TeTriRF: Temporal Tri-Plane Radiance Fields for Efficient Free-Viewpoint Video. 6487-6496
- Jaskirat Singh, Jianming Zhang, Qing Liu, Cameron Smith, Zhe Lin, Liang Zheng:
SmartMask: Context Aware High-Fidelity Mask Generation for Fine-grained Object Insertion and Layout Control. 6497-6506
- Ozgur Kara, Bariscan Kurtkaya, Hidir Yesiltepe, James M. Rehg, Pinar Yanardag:
RAVE: Randomized Noise Shuffling for Fast and Consistent Video Editing with Diffusion Models. 6507-6516
- Yixun Liang, Xin Yang, Jiantao Lin, Haodong Li, Xiaogang Xu, Yingcong Chen:
LucidDreamer: Towards High-Fidelity Text-to-3D Generation via Interval Score Matching. 6517-6526
- Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Wei Wei, Tingbo Hou, Yael Pritch, Neal Wadhwa, Michael Rubinstein, Kfir Aberman:
HyperDreamBooth: HyperNetworks for Fast Personalization of Text-to-Image Models. 6527-6536
- Yujie Wei, Shiwei Zhang, Zhiwu Qing, Hangjie Yuan, Zhiheng Liu, Yu Liu, Yingya Zhang, Jingren Zhou, Hongming Shan:
Dream Video: Composing Your Dream Videos with Customized Subject and Motion. 6537-6549
- Tao Hu, Fangzhou Hong, Ziwei Liu:
SurMo: Surface-based 4D Motion Modeling for Dynamic Human Rendering. 6550-6560
- Tomás Soucek, Dima Damen, Michael Wray, Ivan Laptev, Josef Sivic:
GenHowTo: Learning to Generate Actions and State Transformations from Instructional Videos. 6561-6571
- Xiang Wang, Shiwei Zhang, Hangjie Yuan, Zhiwu Qing, Biao Gong, Yingya Zhang, Yujun Shen, Changxin Gao, Nong Sang:
A Recipe for Scaling up Text-to-Video Generation with Text-free Videos. 6572-6582
- Yunqi Miao, Jiankang Deng, Jungong Han:
WaveFace: Authentic Face Restoration with Efficient Frequency Recovery. 6583-6592
- Xi Chen, Lianghua Huang, Yu Liu, Yujun Shen, Deli Zhao, Hengshuang Zhao:
AnyDoor: Zero-shot Object-level Image Customization. 6593-6602
- Moayed Haji-Ali, Guha Balakrishnan, Vicente Ordonez:
ElasticDiffusion: Training-Free Arbitrary Size Image Generation Through Global-Local Content Separation. 6603-6612
- Tianwei Yin, Michaël Gharbi, Richard Zhang, Eli Shechtman, Frédo Durand, William T. Freeman, Taesung Park:
One-Step Diffusion with Distribution Matching Distillation. 6613-6623
- Biao Gong, Siteng Huang, Yutong Feng, Shiwei Zhang, Yuyuan Li, Yu Liu:
Check, Locate, Rectify: A Training-Free Layout Calibration System for Text- to- Image Generation. 6624-6634
- Zhiwu Qing, Shiwei Zhang, Jiayu Wang, Xiang Wang, Yujie Wei, Yingya Zhang, Changxin Gao, Nong Sang:
Hierarchical Spatio-temporal Decoupling for Text-to- Video Generation. 6635-6645

- Xian Liu, Xiaohang Zhan, Jiaxiang Tang, Ying Shan, Gang Zeng, Dahua Lin, Xihui Liu, Ziwei Liu: **HumanGaussian: Text-Driven 3D Human Generation with Gaussian Splatting.** 6646-6657

Hong-Xing Yu, Haoyi Duan, Junhwa Hur, Kyle Sargent, Michael Rubinstein, William T. Freeman, Forrester Cole, Deqing Sun, Noah Snavely, Jiajun Wu, Charles Herrmann: **WonderJourney: Going from Anywhere to Everywhere.** 6658-6667

Rishabh Parihar, Abhijnya Bhat, Abhipsa Basu, Saswat Mallick, Jogendra Nath Kundu, R. Venkatesh Babu: **Balancing Act: Distribution-Guided Debiasing in Diffusion Models.** 6668-6678

Jan-Niklas Dihlmann, Andreas Engelhardt, Hendrik P. A. Lensch: **SIGNeRF: Scene Integrated Generation for Neural Radiance Fields.** 6679-6688

Yuming Jiang, Tianxing Wu, Shuai Yang, Chenyang Si, Dahua Lin, Yu Qiao, Chen Change Loy, Ziwei Liu: **VideoBooth: Diffusion-based Video Generation with Image Prompts.** 6689-6700

Bowei Chen, Brian Curless, Ira Kemelmacher-Shlizerman, Steven M. Seitz: **Total Selfie: Generating Full-Body Selfies.** 6701-6711

Ruoyu Feng, Wenming Weng, Yanhui Wang, Yuhui Yuan, Jianmin Bao, Chong Luo, Zhibo Chen, Baining Guo: **CCEdit: Creative and Controllable Video Editing via Diffusion Models.** 6712-6722

Xuekun Jiang, Anyi Rao, Jingbo Wang, Dahua Lin, Bo Dai: **Cinematic Behavior Transfer via NeRF-based Differentiable Filming.** 6723-6732

Kelvin C. K. Chan, Yang Zhao, Xuhui Jia, Ming-Hsuan Yang, Huisheng Wang: **Improving Subject-Driven Image Synthesis with Subject-Agnostic Guidance.** 6733-6742

Haofeng Liu, Chenshu Xu, Yifei Yang, Lihua Zeng, Shengfeng He: **Drag Your Noise: Interactive Point-based Editing via Diffusion Semantic Propagation.** 6743-6752

Ta Ying Cheng, Matheus Gadelha, Thibault Groueix, Matthew Fisher, Radomír Mech, Andrew Markham, Niki Trigoni: **Learning Continuous 3D Words for Text-to-Image Generation.** 6753-6762

Yao Ni, Piotr Koniusz: **\$\bigcirc\!\!\$\texttt{CHAIN}: Enhancing Generalization in Data-Efficient GANs via LipsCHitz Continuity ConstrAlned Normalization.** 6763-6774

Jeong-gi Kwak, Erqun Dong, Yuhe Jin, Hanseok Ko, Shweta Mahajan, Kwang Moo Yi: **ViViD-1-to-3: Novel View Synthesis with Video Diffusion Models.** 6775-6785

Yu Zeng, Vishal M. Patel, Haochen Wang, Xun Huang, Ting-Chun Wang, Ming-Yu Liu, Yogesh Balaji: **JeDi: Joint-Image Diffusion Models for Finetuning-Free Personalized Text-to-Image Generation.** 6786-6795

Taoran Yi, Jiemin Fang, Junjie Wang, Guanjun Wu, Lingxi Xie, Xiaopeng Zhang, Wenyu Liu, Qi Tian, Xinggang Wang: **GaussianDreamer: Fast Generation from Text to 3D Gaussians by Bridging 2D and 3D Diffusion Models.** 6796-6807

Shweta Mahajan, Tanzila Rahman, Kwang Moo Yi, Leonid Sigal: **Prompting Hard or Hardly Prompting: Prompt Inversion for Text-to-Image Diffusion Models.** 6808-6817

Dewei Zhou, You Li, Fan Ma, Xiaoting Zhang, Yi Yang: **MIGC: Multi-Instance Generation Controller for Text-to-Image Synthesis.** 6818-6828

Qihang Zhang, Chaoyang Wang, Aliaksandr Siarohin, Peiye Zhuang, Yinghao Xu, Ceyuan Yang, Dahua Lin, Bolei Zhou, Sergey Tulyakov, Hsin-Ying Lee: **Towards Text-guided 3D Scene Composition.** 6829-6838

Qihang Zhang, Yinghao Xu, Yujun Shen, Bo Dai, Bolei Zhou, Ceyuan Yang: **BerfScene: Bev-conditioned Equivariant Radiance Fields for Infinite 3D Scene Generation.** 6839-6849

Kaede Shiohara, Toshihiko Yamasaki: **Face2Diffusion for Fast and Editable Face Personalization.** 6850-6859

Pengyang Ling, Lin Chen, Pan Zhang, Huaian Chen, Yi Jin, Jinjin Zheng: **FreeDrag: Feature Dragging for Reliable Point-Based Image Editing.** 6860-6870

Dongyoung Choi, Hyeonjoong Jang, Min H. Kim: **OmniLocalRF: Omnidirectional Local Radiance Fields from Dynamic Videos.** 6871-6880

Qihao Liu, Yi Zhang, Song Bai, Adam Kortylewski, Alan L. Yuille: **DIRECT-3D: Learning Direct Text-to-3D Generation on Massive Noisy 3D Data.** 6881-6891

Bin Fu, Fanghua Yu, Anran Liu, Zixuan Wang, Jie Wen, Junjun He, Yu Qiao: **Generate Like Experts: Multi-Stage Font Generation by Incorporating Font Transfer Process into Diffusion Models.** 6892-6901

Yuqing Wen, Yucheng Zhao, Yingfei Liu, Fan Jia, Yanhui Wang, Chong Luo, Chi Zhang, Tiancai Wang, Xiaoyan Sun, Xiangyu Zhang: **Panacea: Panoramic and Controllable Video Generation for Autonomous Driving.** 6902-6912

Qian Wang, Weiqi Li, Chong Mou, Xinhua Cheng, Jian Zhang: **360DVD: Controllable Panorama Video Generation with 360-Degree Video Diffusion Model.** 6913-6923

Mehdi Safaei, Aryan Mikaeili, Or Patashnik, Daniel Cohen-Or, Ali Mahdavi-Amiri: **CLiC: Concept Learning in Context.** 6924-6933



- Yingying Deng, Xiangyu He, Fan Tang, Weiming Dong:
Z*: Zero-shot Style Transfer via Attention Reweighting. 6934-6944
- Pengze Zhang, Hubery Yin, Chen Li, Xiaohua Xie:
Tackling the Singularities at the Endpoints of Time Intervals in Diffusion Models. 6945-6954
- Shikai Li, Jianglin Fu, Kaiyuan Liu, Wentao Wang, Kwan-Yee Lin, Wayne Wu:
CosmicMan: A Text-to-Image Foundation Model for Humans. 6955-6965
- Runze He, Shaofei Huang, Xuecheng Nie, Tianrui Hui, Luoqi Liu, Jiao Dai, Jizhong Han, Guanbin Li, Si Liu:
Customize your NeRF: Adaptive Source Driven 3D Scene Editing via Local-Global Iterative Training. 6966-6975
- Shuliang Ning, Duomin Wang, Yipeng Qin, Zirong Jin, Baoyuan Wang, Xiaoguang Han:
PICTURE: Photorealistic Virtual Try-on from Unconstrained dEsigns. 6976-6985
- Qin Guo, Tianwei Lin:
Focus on Your Instruction: Fine-grained and Multi-instruction Image Editing by Attention Modulation. 6986-6996
- Ziying Huang, Fan Tang, Yong Zhang, Xiaodong Cun, Juan Cao, Jintao Li, Tong-Yee Lee:
Make-Your-Anchor: A Diffusion-based 2D Avatar Generation Framework. 6997-7006
- Zanlin Ni, Yulin Wang, Renping Zhou, Jiayi Guo, Jinyi Hu, Zhiyuan Liu, Shiji Song, Yuan Yao, Gao Huang:
Revisiting Non-Autoregressive Transformers for Efficient Image Synthesis. 7007-7016
- Xu Yang, Changxing Ding, Zhibin Hong, Junhao Huang, Jin Tao, Xiangmin Xu:
Texture-Preserving Diffusion Models for High-Fidelity Virtual Try-On. 7017-7026
- Junyi Yao, Yijiang Liu, Zhen Dong, Mingfei Guo, Helan Hu, Kurt Keutzer, Li Du, Daquan Zhou, Shanghang Zhang:
PromptCoT: Align Prompt Distribution via Adapted Chain-of-Thought. 7027-7037
- Willi Menapace, Aliaksandr Siarohin, Ivan Skorokhodov, Ekaterina Deyneka, Tsai-Shien Chen, Anil Kag, Yuwei Fang, Aleksei Stolar, Elisa Ricci, Jian Ren, Sergey Tulyakov:
Snap Video: Scaled Spatiotemporal Transformers for Text-to-Video Synthesis. 7038-7048
- Zhipeng Cai, Matthias Mueller, Reiner Birk, Diana Wofk, Shao-Yen Tseng, Junda Cheng, Gabriela Ben Melech Stan, Vasudev Lal, Michael Paulitsch:
L-MAGIC: Language Model Assisted Generation of Images with Coherence. 7049-7058
- Yuanze Lin, Yi-Wen Chen, Yi-Hsuan Tsai, Lu Jiang, Ming-Hsuan Yang:
Text-Driven Image Editing via Learnable Regions. 7059-7068
- Seongmin Hong, Kyeonghyun Lee, Suh Yoon Jeon, Hyewon Bae, Se Young Chun:
On Exact Inversion of DPM-Solvers. 7069-7078
- Jiayu Yang, Ziang Cheng, Yunfei Duan, Pan Ji, Hongdong Li:
ConsistNet: Enforcing 3D Consistency for Multi-View Images Diffusion. 7079-7088
- Ruiqi Wu, Liangyu Chen, Tong Yang, Chunlei Guo, Chongyang Li, Xiangyu Zhang:
LAMP: Learn A Motion Pattern for Few-Shot Video Generation. 7089-7098
- Pengfei Zhu, Yang Sun, Bing Cao, Qinghua Hu:
Task-Customized Mixture of Adapters for General Image Fusion. 7099-7108
- Yuyang Yu, Bangzhen Liu, Chenxi Zheng, Xuemiao Xu, Shengfeng He, Huaidong Zhang:
Beyond Textual Constraints: Learning Novel Diffusion Conditions with Fewer Examples. 7109-7118
- Yu Deng, Duomin Wang, Xiaohang Ren, Xingyu Chen, Baoyuan Wang:
Portrait4D: Learning One-Shot 4D Head Avatar Synthesis using Synthetic Data. 7119-7130
- Dengsheng Chen, Xiaoming Wei, Xiaolin Wei:
Animating General Image with Large Visual Motion Model. 7131-7140
- Zuoyue Li, Zhenqiang Li, Zhaopeng Cui, Marc Pollefeys, Martin R. Oswald:
Sat2Scene: 3D Urban Scene Generation from Satellite Images with Diffusion. 7141-7150
- Yazhou Xing, Yingqing He, Zeyue Tian, Xintao Wang, Qifeng Chen:
Seeing and Hearing: Open-domain Visual-Audio Generation with Diffusion Latent Aligners. 7151-7161
- Zhixing Zhang, Bichen Wu, Xiaoyan Wang, Yaqiao Luo, Luxin Zhang, Yinan Zhao, Peter Vajda, Dimitris N. Metaxas, Licheng Yu:
AVID: Any-Length Video Inpainting with Diffusion Model. 7162-7172
- Xiaojuan Wang, Janne Kontkanen, Brian Curless, Steven M. Seitz, Ira Kemelmacher-Shlizerman, Ben Mildenhall, Pratul P. Srinivasan, Dor Verbin, Aleksander Holynski:
Generative Powers of Ten. 7173-7182
- Muyang Li, Tianle Cai, Jiaxin Cao, Qinsheng Zhang, Han Cai, Junjie Bai, Yangqing Jia, Kai Li, Song Han:
DistriFusion: Distributed Parallel Inference for High-Resolution Diffusion Models. 7183-7193
- Han Cai, Muyang Li, Qinsheng Zhang, Ming-Yu Liu, Song Han:
Condition-Aware Neural Network for Controlled Image Generation. 7194-7203
- Subhadeep Koley, Ayan Kumar Bhunia, Deependra Sekhri, Aneeshan Sain, Pinaki Nath Chowdhury, Tao Xiang, Yi-Zhe Song:
It's All About Your Sketch: Democratizing Sketch Control in Diffusion Models. 7204-7214
- Pengchong Qiao, Lei Shang, Chang Liu, Baigui Sun, Xiangyang Ji, Jie Chen:
FaceChain-SuDe: Building Derived Class to Inherit Category Attributes for One-Shot Subject-Driven Generation.



7215-7224

- Yiran Xu, Zhixin Shu, Cameron Smith, Seoung Wug Oh, Jia-Bin Huang:
In-N-Out: Faithful 3D GAN Inversion with Volumetric Decomposition for Face Editing. 7225-7235
- Gaurav Shrivastava, Abhinav Shrivastava:
Video Prediction by Modeling Videos as Continuous Multi-Dimensional Processes. 7236-7245
- Yibo Wang, Ruiyuan Gao, Kai Chen, Kaiqiang Zhou, Yingjie Cai, Lanqing Hong, Zhenguo Li, Lihui Jiang, Dit-Yan Yeung, Qiang Xu, Kai Zhang:
DetDiffusion: Synergizing Generative and Perceptive Models for Enhanced Data Generation and Perception. 7246-7255
- Ling Yang, Haotian Qian, Zhilong Zhang, Jingwei Liu, Bin Cui:
Structure-Guided Adversarial Training of Diffusion Models. 7256-7266
- Tianshui Chen, Jianman Lin, Zhijing Yang, Chunmei Qing, Liang Lin:
Learning Adaptive Spatial Coherent Correlations for Speech-Preserving Facial Expression Manipulation. 7267-7276
- Songwei Ge, Aniruddha Mahapatra, Gaurav Parmar, Jun-Yan Zhu, Jia-Bin Huang:
On the Content Bias in Fréchet Video Distance. 7277-7288
- Junyu Zhang, Daochang Liu, Eunbyung Park, Shichao Zhang, Chang Xu:
Residual Learning in Diffusion Models. 7289-7299
- Yufeng Zheng, Xuetong Li, Koki Nagano, Sifei Liu, Otmar Hilliges, Shalini De Mello:
A Unified Approach for Text-and Image-Guided 4D Scene Generation. 7300-7309
- Haoxin Chen, Yong Zhang, Xiaodong Cun, Menghan Xia, Xintao Wang, Chao Weng, Ying Shan:
VideoCrafter2: Overcoming Data Limitations for High-Quality Video Diffusion Models. 7310-7320
- Guilherme G. Schardong, Tiago Novello, Hallison Paz, Iurii Medvedev, Vinícius da Silva, Luiz Velho, Nuno Gonçalves:
Neural Implicit Morphing of Face Images. 7321-7330
- Minghui Hu, Jianbin Zheng, Chuanxia Zheng, Chaoyue Wang, Dacheng Tao, Tat-Jen Cham:
One More Step: A Versatile Plug-and-Play Module for Rectifying Diffusion Schedule Flaws and Enhancing Low-Frequency Controls. 7331-7340
- Siddhant Jain, Daniel Watson, Eric Tabellion, Aleksander Holynski, Ben Poole, Janne Kontkanen:
Video Interpolation with Diffusion Models. 7341-7351
- Junming Chen, Yunfei Liu, Jianan Wang, Ailing Zeng, Yu Li, Qifeng Chen:
DiffSHEG: A Diffusion-Based Approach for Real-Time Speech-Driven Holistic 3D Expression and Gesture Generation. 7352-7361
- Yushi Huang, Ruihao Gong, Jing Liu, Tianlong Chen, Xianglong Liu:
TFMQ-DM: Temporal Feature Maintenance Quantization for Diffusion Models. 7362-7371
- Huijie Zhang, Yifu Lu, Ismail Alkhouri, Saiprasad Ravishankar, Dogyoon Song, Qing Qu:
Improving Training Efficiency of Diffusion Models via Multi-Stage Framework and Tailored Multi-Decoder Architecture. 7372-7381
- Lijie Fan, Kaifeng Chen, Dilip Krishnan, Dina Katabi, Phillip Isola, Yonglong Tian:
Scaling Laws of Synthetic Images for Model Training ... for Now. 7382-7392
- Fengyuan Shi, Jiaxi Gu, Hang Xu, Songcen Xu, Wei Zhang, Limin Wang:
BIVDiff: A Training-Free Framework for General-Purpose Video Synthesis via Bridging Image and Video Diffusion Models. 7393-7402
- Haoyu Ma, Shahin Mahdizadehaghdam, Bichen Wu, Zhipeng Fan, Yuchao Gu, Wenliang Zhao, Lior Shapira, Xiaohui Xie:
MaskINT: Video Editing via Interpolative Non-autoregressive Masked Transformers. 7403-7412
- Gee-Sern Jason Hsu, Jie-Ying Zhang, Huang Yu Hsiang, Wei-Jie Hong:
Pose Adapted Shape Learning for Large-Pose Face Reenactment. 7413-7422
- Fei Deng, Qifei Wang, Wei Wei, Tingbo Hou, Matthias Grundmann:
PRDP: Proximal Reward Difference Prediction for Large-Scale Reward Finetuning of Diffusion Models. 7423-7433
- Leigang Qu, Wenjie Wang, Yongqi Li, Hanwang Zhang, Liqiang Nie, Tat-Seng Chua:
Discriminative Probing and Tuning for Text-to-Image Generation. 7434-7444
- Dawit Mureja Argaw, Mattia Soldan , Alejandro Pardo, Chen Zhao, Fabian Caba Heilbron, Joon Son Chung, Bernard Ghanem :
Towards Automated Movie Trailer Generation. 7445-7454
- Qingguo Liu , Chenyi Zhuang, Pan Gao, Jie Qin:
CDFormer: When Degradation Prediction Embraces Diffusion Model for Blind Image Super-Resolution. 7455-7464
- Sicheng Mo, Fangzhou Mu, Kuan Heng Lin, Yanli Liu, Bochen Guan, Yin Li, Bolei Zhou:
FreeControl: Training-Free Spatial Control of Any Text-to-Image Diffusion Model with Any Condition. 7465-7475
- Mengqi Huang, Zhendong Mao, Mingcong Liu, Qian He, Yongdong Zhang:
RealCustom: Narrowing Real Text Word for Real-Time Open-Domain Text-to-Image Customization. 7476-7485
- Xirui Li, Chao Ma, Xiaokang Yang, Ming-Hsuan Yang:
VidToMe: Video Token Merging for Zero-Shot Video Editing. 7486-7495



- Qilong Zhangli, Jindong Jiang, Di Liu, Licheng Yu, Xiaoliang Dai, Ankit Ramchandani, Guan Pang, Dimitris N. Metaxas, Praveen Krishnan:
Layout-Agnostic Scene Text Image Synthesis with Diffusion Models. 7496-7506
- Zhan Peng, Xinyi Ye, Weiyue Zhao, Tianqi Liu, Huijiang Sun, Baopu Li, Zhiguo Cao:
3D Multi-frame Fusion for Video Stabilization. 7507-7516
- Huijiang Sun, Xingyi Li, Liao Shen, Xinyi Ye, Ke Xian, Zhiguo Cao:
DyBluRF: Dynamic Neural Radiance Fields from Blurry Monocular Video. 7517-7527
- Maomao Li, Yu Li, Tianyu Yang, Yunfei Liu, Dongxu Yue, Zhihui Lin, Dong Xu:
A Video is Worth 256 Bases: Spatial-Temporal Expectation-Maximization Inversion for Zero-Shot Video Editing. 7528-7537
- Xiao-Juan Li, Dingxi Zhang, Shu-Yu Chen, Feng-Lin Liu:
StrokeFaceNeRF: Stroke-Based Facial Appearance Editing in Neural Radiance Field. 7538-7547
- Jiayi Guo, Xinqian Xu, Yifan Pu, Zanlin Ni, Chaofei Wang, Manushree Vasu, Shiji Song, Gao Huang, Humphrey Shi:
Smooth Diffusion: Crafting Smooth Latent Spaces in Diffusion Models. 7548-7558
- Mengyao Lyu, Yuhong Yang, Haiwen Hong, Hui Chen, Xuan Jin, Yuan He, Hui Xue, Jungong Han, Guiguang Ding:
One-dimensional Adapter to Rule Them All: Concepts, Diffusion Models and Erasing Applications. 7559-7568
- Ivan Skorokhodov, Willi Menapace, Aliaksandr Siarohin, Sergey Tulyakov:
Hierarchical Patch Diffusion Models for High-Resolution Video Generation. 7569-7579
- Saeed Khorram, Mingqi Jiang, Mohamad Shahbazi, Mohamad H. Danesh, Fuxin Li:
Taming the Tail in Class-Conditional GANs: Knowledge Sharing via Unconditional Training at Lower Resolutions. 7580-7590
- Haiwei Chen, Yajie Zhao:
Don't Look into the Dark: Latent Codes for Pluralistic Image Inpainting. 7591-7600
- Zhaoyang Sun, Shengwu Xiong, Yaxiong Chen, Yi Rong:
Content-Style Decoupling for Unsupervised Makeup Transfer without Generating Pseudo Ground Truth. 7601-7610
- Shengqu Cai, Duygu Ceylan, Matheus Gadelha, Chun-Hao Paul Huang, Tuanfeng Yang Wang, Gordon Wetzstein:
Generative Rendering: Controllable 4D-Guided Video Generation with 2D Diffusion Models. 7611-7620
- Yuchao Gu, Yipin Zhou, Bichen Wu, Licheng Yu, Jia-Wei Liu, Rui Zhao, Jay Zhangjie Wu, David Junhao Zhang, Mike Zheng Shou, Kevin Tang:
VideoSwap: Customized Video Subject Swapping with Interactive Semantic Point Correspondence. 7621-7630
- Yuchao Gu, Xintao Wang, Yixiao Ge, Ying Shan, Mike Zheng Shou:
Rethinking the Objectives of Vector-Quantized Tokenizers for Image Synthesis. 7631-7640
- Hao Fei, Shengqiong Wu, Wei Ji, Hanwang Zhang, Tat-Seng Chua:
Dysen-VDM: Empowering Dynamics-Aware Text-to-Video Diffusion with LLMs. 7641-7653
- Tianqi Liu, Xinyi Ye, Min Shi, Zihao Huang, Zhiyu Pan, Zhan Peng, Zhiguo Cao:
Geometry-aware Reconstruction and Fusion-refined Rendering for Generalizable Neural Radiance Fields. 7654-7663
- Jia-Wei Liu, Yan-Pei Cao, Jay Zhangjie Wu, Weijia Mao, Yuchao Gu, Rui Zhao, Jussi Keppo, Ying Shan, Mike Zheng Shou:
DynVideo-E: Harnessing Dynamic NeRF for Large-Scale Motion- and View-Change Human-Centric Video Editing. 7664-7674
- Yibin Wang, Weizhong Zhang, Jianwei Zheng, Cheng Jin:
High-fidelity Person-centric Subject-to-Image Synthesis. 7675-7684
- Yinwei Wu, Xingyi Yang, Xinchao Wang:
Relation Rectification in Diffusion Model. 7685-7694
- Karran Pandey, Paul Guerrero, Matheus Gadelha, Yannick Hold-Geoffroy, Karan Singh, Niloy J. Mitra:
Diffusion Handles Enabling 3D Edits for Diffusion Models by Lifting Activations to 3D. 7695-7704
- Chenjie Cao, Yunuo Cai, Qiaole Dong, Yikai Wang, Yanwei Fu:
LeftRefill: Filling Right Canvas based on Left Reference through Generalized Text-to-Image Diffusion Model. 7705-7715
- Andre Rochow, Max Schwarz, Sven Behnke:
FSRT: Facial Scene Representation Transformer for Face Reenactment from Factorized Appearance, Head-Pose, and Facial Expression Features. 7716-7726
- Zijie Chen, Lichao Zhang, Fangsheng Weng, Lili Pan, Zhenzhong Lan:
Tailored Visions: Enhancing Text-to-Image Generation with Personalized Prompt Rewriting. 7727-7736
- Yijun Yang, Ruiyuan Gao, Xiaosen Wang, Tsung-Yi Ho, Nan Xu, Qiang Xu:
MMA-Diffusion: MultiModal Attack on Diffusion Models. 7737-7746
- Yiming Zhang, Zhening Xing, Yanhong Zeng, Youqing Fang, Kai Chen:
PIA: Your Personalized Image Animator via Plug-and-Play Modules in Text-to-Image Models. 7747-7756
- Baoquan Zhang, Huaibin Wang, Chuyao Luo, Xutao Li, Guotao Liang, Yunming Ye, Xiaochen Qi, Yao He:
Codebook Transfer with Part-of-Speech for Vector-Quantized Image Modeling. 7757-7766



- Yang Zhou, Rongjun Xiao, Dani Lischinski, Daniel Cohen-Or, Hui Huang:
Generating Non-Stationary Textures Using Self-Rectification. 7767-7776
- Zhenyu Zhou, Defang Chen, Can Wang, Chun Chen:
Fast ODE-based Sampling for Diffusion Models in Around 5 Steps. 7777-7786
- Yang Zhou, Zichong Chen, Hui Huang:
Deformable One-Shot Face Stylistization via DINO Semantic Guidance. 7787-7796
- Siteng Huang, Biao Gong, Yutong Feng, Xi Chen, Yuqian Fu, Yu Liu, Donglin Wang:
Learning Disentangled Identifiers for Action-Customized Text-to-Image Generation. 7797-7806
- Thuan Hoang Nguyen, Anh Tran:
SwiftBrush: One-Step Text-to-Image Diffusion Model with Variational Score Distillation. 7807-7816
- Bingyan Liu, Chengyu Wang, Tingfeng Cao, Kui Jia, Jun Huang:
Towards Understanding Cross and Self-Attention in Stable Diffusion for Text-Guided Image Editing. 7817-7826
- Zhen Xing, Qi Dai, Han Hu, Zuxuan Wu, Yu-Gang Jiang:
SimDA: Simple Diffusion Adapter for Efficient Video Generation. 7827-7839
- Tariq Berrada, Jakob Verbeek, Camille Couprie, Kartek Alahari:
Unlocking Pre-Trained Image Backbones for Semantic Image Synthesis. 7840-7849
- Hang Yu, Ruilin Li, Shaorong Xie, Jiayan Qiu:
Shadow-Enlightened Image Outpainting. 7850-7860
- Hsin-Ying Lee, Hung-Yu Tseng, Hsin-Ying Lee, Ming-Hsuan Yang:
Exploiting Diffusion Prior for Generalizable Dense Prediction. 7861-7871
- Jongwoo Choi, Kwanggyo Seo, Amirsaman Ashtari, Junyong Noh:
StyleCineGAN: Landscape Cinemagraph Generation Using a Pre-trained StyleGAN. 7872-7881
- Shuyuan Tu, Qi Dai, Zhi-Qi Cheng, Han Hu, Xintong Han, Zuxuan Wu, Yu-Gang Jiang:
MotionEditor: Editing Video Motion via Content-Aware Diffusion. 7882-7891
- Zixuan Wang, Jia Jia, Shikun Sun, Haozhe Wu, Rong Han, Zhenyu Li, Di Tang, Jiaqing Zhou, Jiebo Luo:
DanceCamera3D: 3D Camera Movement Synthesis with Music and Dance. 7892-7901
- Jiwoo Chung, Sangeek Hyun, Sang-Heon Shim, Jae-Pil Heo:
Diversity-Aware Channel Pruning for StyleGAN Compression. 7902-7911
- Kaiwen Zhang, Yifan Zhou, Xudong Xu, Bo Dai, Xingang Pan:
DiffMorpher: Unleashing the Capability of Diffusion Models for Image Morphing. 7912-7921
- Sidi Wu, Yizi Chen, Samuel Mermet, Lorenz Hurni, Konrad Schindler, Nicolas Gonthier, Loïc Landrieu:
StegoGAN: Leveraging Steganography for Non-Bijective Image-to-Image Translation. 7922-7931
- Quynh Phung, Songwei Ge, Jia-Bin Huang:
Grounded Text-to-Image Synthesis with Attention Refocusing. 7932-7942
- Vikas Thamizharasan, Difan Liu, Shantanu Agarwal, Matthew Fisher, Michaël Gharbi, Oliver Wang, Alec Jacobson, Evangelos Kalogerakis:
VecFusion: Vector Font Generation with Diffusion. 7943-7952
- Thomas W. Mitchel, Carlos Esteves, Ameesh Makadia:
Single Mesh Diffusion Models with Field Latents for Texture Generation. 7953-7963
- Ryan Po, Guandao Yang, Kfir Aberman, Gordon Wetzstein:
Orthogonal Adaptation for Modular Customization of Diffusion Models. 7964-7973
- Qiqi Hou, Farzad Farhadzadeh, Amir Said, Guillaume Sautière, Hoang Le:
Low-Latency Neural Stereo Streaming. 7974-7984
- Yanyu Li, Xian Liu, Anil Kag, Ju Hu, Yerlan Idelbayev, Dhritiman Sagar, Yanzhi Wang, Sergey Tulyakov, Jian Ren:
TextCraftor: Your Text Encoder can be Image Quality Controller. 7985-7995
- Sherwin Bahmani, Ivan Skorokhodov, Victor Rong, Gordon Wetzstein, Leonidas J. Guibas, Peter Wonka, Sergey Tulyakov, Jeong Joon Park, Andrea Tagliasacchi, David B. Lindell:
4D-fy: Text-to-4D Generation Using Hybrid Score Distillation Sampling. 7996-8006
- Yinbo Chen, Oliver Wang, Richard Zhang, Eli Shechtman, Xiaolong Wang, Michaël Gharbi:
Image Neural Field Diffusion Models. 8007-8017
- Sixian Zhang, Bohan Wang, Junqiang Wu, Yan Li, Tingting Gao, Di Zhang, Zhongyuan Wang:
Learning Multi-Dimensional Human Preference for Text-to-Image Generation. 8018-8027
- Tingting Zheng, Kui Jiang, Hongxun Yao:
Dynamic Policy-Driven Adaptive Multi-Instance Learning for Whole Slide Image Classification. 8028-8037
- Haipeng Liu, Yang Wang, Biao Qian, Meng Wang, Yong Rui:
Structure Matters: Tackling the Semantic Discrepancy in Diffusion Models for Image Inpainting. 8038-8047
- Yizhi Song, Zhifei Zhang, Zhe Lin, Scott Cohen, Brian L. Price, Jianming Zhang, Soo Ye Kim, He Zhang, Wei Xiong, Daniel G. Aliaga:
IMPRINT: Generative Object Compositing by Learning Identity-Preserving Representation. 8048-8058



- Sizhe Zheng, Pan Gao, Peng Zhou, Jie Qin:
Puff-Net: Efficient Style Transfer with Pure Content and Style Feature Fusion Network. 8059-8068
- Yuxuan Zhang, Yiren Song, Jiaming Liu, Rui Wang, Jinpeng Yu, Hao Tang, Huaxia Li, Xu Tang, Yao Hu, Han Pan, Zhongliang Jing:
SSR-Encoder: Encoding Selective Subject Representation for Subject-Driven Generation. 8069-8078
- Yash Jain, Anshul Nasery, Vibhav Vineet, Harkirat S. Behl:
Peekaboo: Interactive Video Generation via Masked-Diffusion. 8079-8088
- Hao Ouyang, Qiuyu Wang, Yuxi Xiao, Qingyan Bai, Juntao Zhang, Kecheng Zheng, Xiaowei Zhou, Qifeng Chen, Qifeng Chen:
CoDeF: Content Deformation Fields for Temporally Consistent Video Processing. 8089-8099
- Jisu Nam, Heesu Kim, Dongjae Lee, Siyoon Jin, Seungryong Kim, Seunggyu Chang:
DreamMatcher: Appearance Matching Self-Attention for Semantically-Consistent Text-to-Image Personalization. 8100-8110
- Yunhan Yang, Yukun Huang, Xiaoyang Wu, Yuan-Chen Guo, Song-Hai Zhang, Hengshuang Zhao, Tong He, Xihui Liu:
DreamComposer: Controllable 3D Object Generation via Multi-View Conditions. 8111-8120
- Qingyang Liu, Junqi You, Jianting Wang, Xinhao Tao, Bo Zhang, Li Niu:
Shadow Generation for Composite Image Using Diffusion Model. 8121-8130
- Min Wei, Jingkai Zhou, Junyao Sun, Xuesong Zhang:
Adversarial Score Distillation: When Score Distillation Meets GAN. 8131-8141
- Yuang Ai, Xiaoqiang Zhou, Huaibo Huang, Lei Zhang, Ran He:
Uncertainty-Aware Source-Free Adaptive Image Super-Resolution with Wavelet Augmentation Transformer. 8142-8152
- Li Hu:
Animate Anyone: Consistent and Controllable Image-to-Video Synthesis for Character Animation. 8153-8163
- ChangHee Yang, Chanhee Kang, Kyeongbo Kong, Hanni Oh, Suk-Ju Kang:
Person in Place: Generating Associative Skeleton-Guidance Maps for Human-Object Interaction Image Editing. 8164-8175
- Jeongho Kim, Gyojung Gu, Minho Park, Sunghyun Park, Jaegul Choo:
Stable VITON: Learning Semantic Correspondence with Latent Diffusion Model for Virtual Try-On. 8176-8185
- Cusuh Ham, Matthew Fisher, James Hays, Nicholas Kolkin, Yuchen Liu, Richard Zhang, Tobias Hinz:
Personalized Residuals for Concept-Driven Text-to-Image Generation. 8186-8195
- Yanwu Xu, Yang Zhao, Zhisheng Xiao, Tingbo Hou:
UFOGen: You Forward Once Large Scale Text-to-Image Generation via Diffusion GANs. 8196-8206
- Feng Liang, Bichen Wu, Jialiang Wang, Licheng Yu, Kunpeng Li, Yinan Zhao, Ishan Misra, Jia-Bin Huang, Peizhao Zhang, Peter Vajda, Diana Marculescu:
FlowVid: Taming Imperfect Optical Flows for Consistent Video-to-Video Synthesis. 8207-8216
- Grace Luo, Trevor Darrell, Oliver Wang, Dan B. Goldman, Aleksander Holynski:
Readout Guidance: Learning Control from Diffusion Features. 8217-8227
- Bram Wallace, Meihua Dang, Rafael Rafailov, Linqi Zhou, Aaron Lou, Senthil Purushwalkam, Stefano Ermon, Caiming Xiong, Shafiq Joty, Nikhil Naik:
Diffusion Model Alignment Using Direct Preference Optimization. 8228-8238
- Jing Nathan Yan, Jiatao Gu, Alexander M. Rush:
Diffusion Models Without Attention. 8239-8249
- Aaron Gokaslan, A. Feder Cooper, Jasmine Collins, Landan Seguin, Austin Jacobson, Mihir Patel, Jonathan Frankle, Cory Stephenson, Volodymyr Kuleshov:
Common Canvas: Open Diffusion Models Trained on Creative-Commons Images. 8250-8260
- Bichen Wu, Ching-Yao Chuang, Xiaoyan Wang, Yichen Jia, Kapil Krishnakumar, Tong Xiao, Feng Liang, Licheng Yu, Peter Vajda:
Fairy: Fast Parallelized Instruction-Guided Video-to-Video Synthesis. 8261-8270
- Thao Nguyen, Utkarsh Ojha, Yuheng Li, Haotian Liu, Yong Jae Lee:
Edit One for All: Interactive Batch Image Editing. 8271-8280
- Chen Zhao, Weiling Cai, Chenyu Dong, Chengwei Hu:
Wavelet-based Fourier Information Interaction with Frequency Diffusion Adjustment for Underwater Image Restoration. 8281-8291
- Shuchen Xue, Zhaoqiang Liu, Fei Chen, Shifeng Zhang, Tianyang Hu, Enze Xie, Zhenguo Li:
Accelerating Diffusion Sampling with Optimized Time Steps. 8292-8301
- Hansam Cho, Jonghyun Lee, Seunggyu Chang, Yonghyun Jeong:
One-Shot Structure-Aware Stylized Image Synthesis. 8302-8311
- Jimyeong Kim, Jungwon Park, Wonjong Rhee:
Selectively Informative Description can Reduce Undesired Embedding Entanglements in Text-to-Image Personalization. 8312-8322



- Junoh Kang, Jinyoung Choi, Sungik Choi, Bohyung Han:
Observation-Guided Diffusion Probabilistic Models. 8323-8331
- Dawit Mureja Argaw, Seunghyun Yoon, Fabian Caba Heilbron, Hanieh Deilamsalehy, Trung Bui, Zhaowen Wang, Franck Dernoncourt, Joon Son Chung:
Scaling Up Video Summarization Pretraining with Large Language Models. 8332-8341
- Jinxin Zhou, Tianyu Ding, Tianyi Chen, Jiachen Jiang, Ilya Zharkov, Zhihui Zhu, Luming Liang:
DREAM: Diffusion Rectification and Estimation-Adaptive Models. 8342-8351
- Amirhossein Habibian, Amir Ghodrati, Noor Fathima, Guillaume Sautière, Risheek Garrepalli, Fatih Porikli, Jens Petersen:
Clockwork Diffusion: Efficient Generation With Model-Step Distillation. 8352-8361
- Yuzhou Huang, Liangbin Xie, Xintao Wang, Ziyang Yuan, Xiaodong Cun, Yixiao Ge, Jiantao Zhou, Chao Dong, Rui Huang, Ruimao Zhang, Ying Shan:
SmartEdit: Exploring Complex Instruction-Based Image Editing with Multimodal Large Language Models. 8362-8371
- Jianhao Zeng, Dan Song, Weizhi Nie, Hongshuo Tian, Tongtong Wang, An-An Liu:
CAT-DM: Controllable Accelerated Virtual Try-On with Diffusion Model. 8372-8382
- Yingbo Zhou, Yutong Ye, Pengyu Zhang, Xian Wei, Mingsong Chen:
Exact Fusion via Feature Distribution Matching for Few-Shot Image Generation. 8383-8392
- Lianyu Pang, Jian Yin, Haoran Xie, Qiping Wang, Qing Li, Xudong Mao:
Cross Initialization for Face Personalization of Text-to-Image Models. 8393-8403
- Xingzhong Hou, Boxiao Liu, Yi Zhang, Jihao Liu, Yu Liu, Haihang You:
EasyDrag: Efficient Point-Based Manipulation on Diffusion Models. 8404-8413
- Yanhui Wang, Jianmin Bao, Wenming Weng, Ruoyu Feng, Dacheng Yin, Tao Yang, Jingxu Zhang, Qi Dai, Zhiyuan Zhao, Chunyu Wang, Kai Qiu, Yuhui Yuan, Xiaoyan Sun, Chong Luo, Baining Guo:
MicroCinema: A Divide-and-Conquer Approach for Text-to-Video Generation. 8414-8424
- Chen Chen, Daochang Liu, Chang Xu:
Towards Memorization-Free Diffusion Models. 8425-8434
- Rui Zhu, Yingwei Pan, Yehao Li, Ting Yao, Zhenglong Sun, Tao Mei, Chang Wen Chen:
SD-DiT: Unleashing the Power of Self-Supervised Discrimination in Diffusion Transformer^{*}. 8435-8445
- Junyan Wang, Zhenhong Sun, Zhiyu Tan, Xuanbai Chen, Weihua Chen, Hao Li, Cheng Zhang, Yang Song:
Towards Effective Usage of Human-Centric Priors in Diffusion Models for Text-based Human Image Generation. 8446-8455
- Guangyang Wu, Xiaohong Liu, Jun Jia, Xuehao Cui, Guangtao Zhai:
Text2QR: Harmonizing Aesthetic Customization and Scanning Robustness for Text-Guided QR Code Generation. 8456-8465
- Danah Yatim, Rafail Fridman, Omer Bar-Tal, Yoni Kasten, Tali Dekel:
Space-Time Diffusion Features for Zero-Shot Text-Driven Motion Transfer. 8466-8476
- Yuhan Liu, Yongjian Deng, Hao Chen, Zhen Yang:
Video Frame Interpolation via Direct Synthesis with the Event-based Reference. 8477-8487
- Chong Mou, Xintao Wang, Jiechong Song, Ying Shan, Jian Zhang:
DiffEditor: Boosting Accuracy and Flexibility on Diffusion-Based Image Editing. 8488-8497
- Nikita Drobyshev, Antoni Bigata Casademunt, Konstantinos Vougioukas, Zoe Landgraf, Stavros Petridis, Maja Pantic:
EMOPortraits: Emotion-Enhanced Multimodal One-Shot Head Avatars. 8498-8507
- Zhan Li, Zhang Chen, Zhong Li, Yi Xu:
Spacetime Gaussian Feature Splatting for Real-Time Dynamic View Synthesis. 8508-8520
- Mengqi Zhang, Yang Fu, Zheng Ding, Sifei Liu, Zhuowen Tu, Xiaolong Wang:
HOIDiffusion: Generating Realistic 3D Hand-Object Interaction Data. 8521-8531
- Alexandros Graikos, Srikanth Yellapragada, Minh-Quan Le, Saarthak Kapse, Prateek Prasanna, Joel H. Saltz, Dimitris Samaras:
Learned Representation-Guided Diffusion Models for Large-Image Generation. 8532-8542
- Jing Shi, Wei Xiong, Zhe Lin, Hyun Joon Jung:
InstantBooth: Personalized Text-to-Image Generation without Test-Time Finetuning. 8543-8552
- Zirui Wang, Zhizhou Sha, Zheng Ding, Yilin Wang, Zhuowen Tu:
TokenCompose: Text-to-Image Diffusion with Token-Level Supervision. 8553-8564
- Hyunyoung Jung, Seonghyeon Nam, Nikolaos Sarafianos, Sungjoo Yoo, Alexander Sorkine-Hornung, Rakesh Ranjan:
Geometry Transfer for Stylizing Radiance Fields. 8565-8575
- Huan Ling, Seung Wook Kim, Antonio Torralba, Sanja Fidler, Karsten Kreis:
Align Your Gaussians: Text-to-4D with Dynamic 3D Gaussians and Composed Diffusion Models. 8576-8588
- Haonan Lin:
DreamSalon: A Staged Diffusion Framework for Preserving Identity-Context in Editable Face Generation. 8589-8598
- Shaoteng Liu, Yuechen Zhang, Wenbo Li, Zhe Lin, Jiaya Jia:
Video-P2P: Video Editing with Cross-Attention Control. 8599-8608



- Vudit Goel, Elia Peruzzo, Yifan Jiang, Dejia Xu, Xingqian Xu, Nicu Sebe, Trevor Darrell, Zhangyang Wang, Humphrey Shi: **PAIR Diffusion: A Comprehensive Multimodal Object-Level Image Editor.** 8609-8618
- Dar-Yen Chen, Hamish Tennent, Ching-Wen Hsu: **ArtAdapter: Text-to-Image Style Transfer using Multi-Level Style Encoder and Explicit Adaptation.** 8619-8628
- Dar-Yen Chen, Ayan Kumar Bhunia, Subhadeep Koley, Aneeshan Sain, Pinaki Nath Chowdhury, Yi-Zhe Song: **DemoCaricature: Democratising Caricature Generation with a Rough Sketch.** 8629-8639
- Zhen Li, Mingdeng Cao, Xintao Wang, Zhongang Qi, Ming-Ming Cheng, Ying Shan: **PhotoMaker: Customizing Realistic Human Photos via Stacked ID Embedding.** 8640-8650
- Kota Sueyoshi, Takashi Matsubara: **Predicated Diffusion: Predicate Logic-Based Attention Guidance for Text-to-Image Diffusion Models.** 8651-8660
- Zhengang Li, Yan Kang, Yuchen Liu, Difan Liu, Tobias Hinz, Feng Liu, Yanzhi Wang: **SNED: Superposition Network Architecture Search for Efficient Video Diffusion Model.** 8661-8670
- Zhongwei Zhang, Fuchen Long, Yingwei Pan, Zhaofan Qiu, Ting Yao, Yang Cao, Tao Mei: **TRIP: Temporal Residual Learning with Image Noise Prior for Image-to-Video Diffusion Models.** 8671-8681
- Xinqian Xu, Jiayi Guo, Zhangyang Wang, Gao Huang, Irfan Essa, Humphrey Shi: **Prompt-Free Diffusion: Taking "Text" Out of Text-to-Image Diffusion Models.** 8682-8692
- Tianhao Qi, Shancheng Fang, Yanze Wu, Hongtao Xie, Jiawei Liu, Lang Chen, Qian He, Yongdong Zhang: **DEADiff: An Efficient Stylization Diffusion Model with Disentangled Representations.** 8693-8702
- Shuai Yang, Yifan Zhou, Ziwei Liu, Chen Change Loy: **Fresco: Spatial-Temporal Correspondence for Zero-Shot Video Translation.** 8703-8712
- Yujian Liu, Yang Zhang, Tommi S. Jaakkola, Shiyu Chang: **Correcting Diffusion Generation Through Resampling.** 8713-8723
- Ruidong Chen, Lanjun Wang, Weizhi Nie, Yongdong Zhang, An-An Liu: **AnyScene: Customized Image Synthesis with Composited Foreground.** 8724-8733
- Taegyeong Lee, Soyeong Kwon, Taehwan Kim: **Grid Diffusion Models for Text-to-Video Generation.** 8734-8743
- Yuanxun Lu, Jingyang Zhang, Shiwei Li, Tian Fang, David McKinnon, Yanghai Tsin, Long Quan, Xun Cao, Yao Yao: **Direct2.5: Diverse Text-to-3D Generation via Multi-view 2.5D Diffusion.** 8744-8753
- Jaehui Hwang, Junghyuk Lee, Jong-Seok Lee: **Anomaly Score: Evaluating Generative Models and Individual Generated Images Based on Complexity and Vulnerability.** 8754-8763
- Marianna Ohanyan, Hayk Manukyan, Zhangyang Wang, Shant Navasardyan, Humphrey Shi: **Zero-Painter: Training-Free Layout Control for Text-to-Image Synthesis.** 8764-8774
- Lingmin Ran, Xiaodong Cun, Jia-Wei Liu, Rui Zhao, Song Zijie, Xintao Wang, Jussi Keppo, Mike Zheng Shou: **X-Adapter: Universal Compatibility of Plugins for Upgraded Diffusion Model.** 8775-8784
- Philipp Schröppel, Christopher Wewer, Jan Eric Lenssen, Eddy Ilg, Thomas Brox: **Neural Point Cloud Diffusion for Disentangled 3D Shape and Appearance Generation.** 8785-8794
- Jiwoo Chung, Sangeek Hyun, Jae-Pil Heo: **Style Injection in Diffusion: A Training-Free Approach for Adapting Large-Scale Diffusion Models for Style Transfer.** 8795-8805
- Shaobin Zhuang, Kunchang Li, Xinyuan Chen, Yaohui Wang, Ziwei Liu, Yu Qiao, Yali Wang: **Vlogger: Make Your Dream A Vlog.** 8806-8817
- Youngjoon Jang, Ji-Hoon Kim, Junseok Ahn, Doyeop Kwak, Hongsun Yang, Yooncheol Ju, Ilhwan Kim, Byeong-Yeol Kim, Joon Son Chung: **Faces that Speak: Jointly Synthesising Talking Face and Speech from Text.** 8818-8828
- Rumeysa Bodur, Binod Bhattacharai, Tae-Kyun Kim: **Prompt Augmentation for Self-supervised Text-guided Image Manipulation.** 8829-8838
- Yujun Shi, Chuhui Xue, Jun Hao Liew, Jiachun Pan, Hanshu Yan, Wenqing Zhang, Vincent Y. F. Tan, Song Bai: **DragDiffusion: Harnessing Diffusion Models for Interactive Point-Based Image Editing.** 8839-8849
- Yan Zeng, Guoqiang Wei, Jian Zheng, Jiaxin Zou, Yang Wei, Yuchen Zhang, Hang Li: **Make Pixels Dance: High-Dynamic Video Generation.** 8850-8860
- Manuel Brack, Felix Friedrich, Katharina Kornmeier, Linoy Tsaban, Patrick Schramowski, Kristian Kersting, Apolinário Passos: **LEDITS++: Limitless Image Editing Using Text-to-Image Models.** 8861-8870
- Shelly Sheynin, Adam Polyak, Uriel Singer, Yuval Kirstain, Amit Zohar, Oron Ashual, Devi Parikh, Yaniv Taigman: **Emu Edit: Precise Image Editing via Recognition and Generation Tasks.** 8871-8879
- Gihyun Kwon, Simon Jenni, Dingzeyu Li, Joon-Young Lee, Jong Chul Ye, Fabian Caba Heilbron: **Concept Weaver: Enabling Multi-Concept Fusion in Text-to-Image Models.** 8880-8889
- Fei Kong, Jinhao Duan, Lichao Sun, Hao Cheng, Renjing Xu, Hengtao Shen, Xiaofeng Zhu, Xiaoshuang Shi, Kaidi Xu: **ACT-Diffusion: Efficient Adversarial Consistency Training for One-Step Diffusion Models.** 8890-8899



- Zicheng Lu, Xiang Guo, Le Hui, Tianrui Chen, Min Yang, Xiao Tang, Feng Zhu, Yuchao Dai:
3D Geometry-aware Deformable Gaussian Splatting for Dynamic View Synthesis. 8900-8910
- Yurui Qian, Qi Cai, Yingwei Pan, Yehao Li, Ting Yao, Qibin Sun, Tao Mei:
Boosting Diffusion Models with Moving Average Sampling in Frequency Domain. 8911-8920
- Takahiro Shirakawa, Seiichi Uchida:
NoiseCollage: A Layout-Aware Text-to-Image Diffusion Model Based on Noise Cropping and Merging. 8921-8930
- Weining Ren, Zihan Zhu, Boyang Sun, Jiaqi Chen, Marc Pollefeys, Songyou Peng:
NeRF On-the-go: Exploiting Uncertainty for Distractor-free NeRFs in the Wild. 8931-8940
- Kai Yang, Jian Tao, Jiafei Lyu, Chunjiang Ge, Jiaxin Chen, Weihan Shen, Xiaolong Zhu, Xiu Li:
Using Human Feedback to Fine-tune Diffusion Models without Any Reward Model. 8941-8951
- Chong Bao, Yinda Zhang, Yuan Li, Xiyu Zhang, Bangbang Yang, Hujun Bao, Marc Pollefeys, Guofeng Zhang, Zhaopeng Cui:
GeneAvatar: Generic Expression-Aware Volumetric Head Avatar Editing from a Single Image. 8952-8963
- Hang Zhang, Anton Savov, Benjamin Dillenburger:
MaskPLAN: Masked Generative Layout Planning from Partial Input. 8964-8973
- Changhoon Kim, Kyle Min, Maitreya Patel, Sheng Cheng, Yezhou Yang:
WOUAF: Weight Modulation for User Attribution and Fingerprinting in Text-to-Image Diffusion Models. 8974-8983
- Zhiyuan Yan, Yuhao Luo, Siwei Lyu, Qingshan Liu, Baoyuan Wu:
Transcending Forgery Specificity with Latent Space Augmentation for Generalizable Deepfake Detection. 8984-8994
- Zeyinzi Jiang, Chaojie Mao, Yulin Pan, Zhen Han, Jingfeng Zhang:
SCEdit: Efficient and Controllable Image Diffusion Generation via Skip Connection Editing. 8995-9004
- Tuna Han Salih Meral, Enis Simsar, Federico Tombari, Pinar Yanardag:
CONFORM: Contrast is All You Need For High-Fidelity Text-to-Image Diffusion Models. 9005-9014
- Haomiao Ni, Bernhard Egger, Suhas Lohit, Anoop Cherian, Ye Wang, Toshiaki Koike-Akino, Sharon X. Huang, Tim K. Marks:
TI2V-Zero: Zero-Shot Image Conditioning for Text-to-Video Diffusion Models. 9015-9025
- Shu Zhang, Xinyi Yang, Yihao Feng, Can Qin, Chia-Chih Chen, Ning Yu, Zeyuan Chen, Huan Wang, Silvio Savarese, Stefano Ermon, Caiming Xiong, Ran Xu:
HIVE: Harnessing Human Feedback for Instructional Visual Editing. 9026-9036
- Peihao Wang, Dejia Xu, Zhiwen Fan, Dilin Wang, Sreyas Mohan, Forrest N. Iandola, Rakesh Ranjan, Yilei Li, Qiang Liu, Zhangyang Wang, Vikas Chandra:
Taming Mode Collapse in Score Distillation for Text-to-3D Generation. 9037-9047
- Kangfu Mei, Mauricio Delbracio, Hossein Talebi, Zhengzhong Tu, Vishal M. Patel, Peyman Milanfar:
CoDi: Conditional Diffusion Distillation for Higher-Fidelity and Faster Image Generation. 9048-9058
- Zakariya Chaouai, Mohamed Tamaazousti:
Universal Robustness via Median Randomized Smoothing for Real-World Super-Resolution. 9059-9068
- Maitreya Patel, Changhoon Kim, Sheng Cheng, Chitta Baral, Yezhou Yang:
ECLIPSE: A Resource-Efficient Text-to-Image Prior for Image Generations. 9069-9078
- Guiwei Zhang, Tianyu Zhang, Guanglin Niu, Zichang Tan, Yalong Bai, Qing Yang:
CAMEL: CAusal Motion Enhancement Tailored for Lifting Text-Driven Video Editing. 9079-9088
- Ganggui Ding, Canyu Zhao, Wen Wang, Zhen Yang, Zide Liu, Hao Chen, Chunhua Shen:
FreeCustom: Tuning-Free Customized Image Generation for Multi-Concept Composition. 9089-9098
- Katherine Xu, Lingzhi Zhang, Jianbo Shi:
Amodal Completion via Progressive Mixed Context Diffusion. 9099-9109
- Zhida Feng, Li Chen, Jing Tian, Jiaxiang Liu, Shikun Feng:
Named Entity Driven Zero-Shot Image Manipulation. 9110-9119
- Lianxin Xie, Csbingbing Zheng, Wen Xue, Le Jiang, Cheng Liu, Si Wu, Hau-San Wong:
Learning Degradation-Unaware Representation with Prior-Based Latent Transformations for Blind Face Restoration. 9120-9129
- Jonas Ricker, Denis Lukovnikov, Asja Fischer:
AEROBLADE: Training-Free Detection of Latent Diffusion Images Using Autoencoder Reconstruction Error. 9130-9140
- Wen Xue, Le Jiang, Lianxin Xie, Si Wu, Yong Xu, Hau-San Wong:
VRetouchEr: Learning Cross-Frame Feature Interdependence with Imperfection Flow for Face Retouching in Videos. 9141-9150
- Juwon Seo, Sung-Hoon Lee, Tae-Young Lee, Seungjun Moon, Gyeong-Moon Park:
Generative Unlearning for Any Identity. 9151-9161
- Xue Song, Jiequan Cui, Hanwang Zhang, Jingjing Chen, Richang Hong, Yu-Gang Jiang:
Doubly Abductive Counterfactual Inference for Text-Based Image Editing. 9162-9171
- Feifan Xu, Rui Li, Si Wu, Yong Xu, Hau-San Wong:
Text-Conditional Attribute Alignment Across Latent Spaces for 3D Controllable Face Image Synthesis. 9172-9181



- Yufan Zhou, Ruiyi Zhang, Jiuxiang Gu, Tong Sun:
Customization Assistant for Text-to-image Generation. 9182-9191
- Hyelin Nam, Gihyun Kwon, Geon Yeong Park, Jong Chul Ye:
Contrastive Denoising Score for Text-Guided Latent Diffusion Image Editing. 9192-9201
- Jinseok Kim, Tae-Kyun Kim:
Arbitrary-Scale Image Generation and Upsampling Using Latent Diffusion Model and Implicit Neural Decoder. 9202-9211
- Hyeonho Jeong, Geon Yeong Park, Jong Chul Ye:
VMC: Video Motion Customization Using Temporal Attention Adaption for Text-to-Video Diffusion Models. 9212-9221
- Mohammad Amin Shabani, Zhaowen Wang, Difan Liu, Nanxuan Zhao, Jimei Yang, Yasutaka Furukawa:
Visual Layout Composer: Image-Vector Dual Diffusion Model for Design Layout Generation. 9222-9231
- Zhikai Chen, Fuchen Long, Zhaofan Qiu, Ting Yao, Wengang Zhou, Jiebo Luo, Tao Mei:
Learning Spatial Adaptation and Temporal Coherence in Diffusion Models for Video Super-Resolution. 9232-9241
- Pablo Marcos-Manchón, Roberto Alcover-Cousó, Juan C. SanMiguel, Jose M. Martínez:
Open-Vocabulary Attention Maps with Token Optimization for Semantic Segmentation in Diffusion Models. 9242-9252
- Jens Eirik Saethre, Roberto Azevedo, Christopher Schroers:
Combining Frame and GOP Embeddings for Neural Video Representation. 9253-9263
- Zhengyao Lv, Yuxiang Wei, Wangmeng Zuo, Kwan-Yee K. Wong:
PLACE: Adaptive Layout-Semantic Fusion for Semantic Image Synthesis. 9264-9274
- Nikita Starodubcev, Dmitry Baranchuk, Artem Fedorov, Artem Babenko:
Your Student is Better than Expected: Adaptive Teacher-Student Collaboration for Text-Conditional Diffusion Models. 9275-9285
- Marco Cannici, Davide Scaramuzza:
Mitigating Motion Blur in Neural Radiance Fields with Events and Frames. 9286-9296
- Yang Yu, Erting Pan, Xinya Wang, Yuheng Wu, Xiaoguang Mei, Jiayi Ma:
Unmixing Before Fusion: A Generalized Paradigm for Multi-Source-Based Hyperspectral Image Synthesis. 9297-9306
- Sadeep Jayasumana, Sri Kumar Ramalingam, Andreas Veit, Daniel Glasner, Ayan Chakrabarti, Sanjiv Kumar:
Rethinking FID: Towards a Better Evaluation Metric for Image Generation. 9307-9315
- Sadeep Jayasumana, Daniel Glasner, Sri Kumar Ramalingam, Andreas Veit, Ayan Chakrabarti, Sanjiv Kumar:
MarkovGen: Structured Prediction for Efficient Text-to-Image Generation. 9316-9325
- Tan Wang, Linjie Li, Kevin Lin, Yuanhao Zhai, Chung-Ching Lin, Zhengyuan Yang, Hanwang Zhang, Zicheng Liu, Lijuan Wang:
Disco: Disentangled Control for Realistic Human Dance Generation. 9326-9336
- Denis Bobkov, Vadim Titov, Aibek Alanov, Dmitry Vetrov:
The Devil is in the Details: StyleFeatureEditor for Detail-Rich StyleGAN Inversion and High Quality Image Editing. 9337-9346
- Hyunjik Kim, Matthias Bauer, Lucas Theis, Jonathan Richard Schwarz, Emilien Dupont:
C3: High-Performance and Low-Complexity Neural Compression from a Single Image or Video. 9347-9358
- Peter Kocsis, Julien Philip, Kalyan Sunkavalli, Matthias Nießner, Yannick Hold-Geoffroy:
LightIt: Illumination Modeling and Control for Diffusion Models. 9359-9369
- Dazhong Shen, Guanglu Song, Zeyue Xue, Fu-Yun Wang, Yu Liu:
Rethinking the Spatial Inconsistency in Classifier-Free Diffusion Guidance. 9370-9379
- Xiefan Guo, Jinlin Liu, Miaomiao Cui, Jiankai Li, Hongyu Yang, Di Huang:
Initno: Boosting Text-to-Image Diffusion Models via Initial Noise Optimization. 9380-9389
- Peng Sun, Bei Shi, Daiwei Yu, Tao Lin:
On the Diversity and Realism of Distilled Dataset: An Efficient Dataset Distillation Paradigm. 9390-9399
- Hao Li, Yang Zou, Ying Wang, Orchid Majumder, Yusheng Xie, R. Manmatha, Ashwin Swaminathan, Zhuowen Tu, Stefano Ermon, Stefano Soatto:
On the Scalability of Diffusion-based Text-to-Image Generation. 9400-9409
- Sanghwan Kim, Hao Tang, Fisher Yu:
Distilling ODE Solvers of Diffusion Models into Smaller Steps. 9410-9419
- Kyle Sargent, Zizhang Li, Tanmay Shah, Charles Herrmann, Hong-Xing Yu, Yunzhi Zhang, Eric Ryan Chan, Dmitry Lagun, Li Fei-Fei, Deqing Sun, Jiajun Wu:
ZeroNVS: Zero-Shot 360-Degree View Synthesis from a Single Image. 9420-9429
- Xingjian Bai, Luke Melas-Kyriazi:
Fixed Point Diffusion Models. 9430-9440
- Rameen Abdal, Yifan Wang, Zifan Shi, Yinghao Xu, Ryan Po, Zhengfei Kuang, Qifeng Chen, Dit-Yan Yeung, Gordon Wetzstein:
Gaussian Shell Maps for Efficient 3D Human Generation. 9441-9451



- Sihan Xu, Yidong Huang, Jiayi Pan, Ziqiao Ma, Joyce Chai:
Inversion-Free Image Editing with Language-Guided Diffusion Models. 9454-9461
- Zhiyuan Ren, Minchul Kim, Feng Liu, Xiaoming Liu:
TIGER: Time-Varying Denoising Model for 3D Point Cloud Generation with Diffusion Process. 9462-9471
- Litu Rout, Yujia Chen, Abhishek Kumar, Constantine Caramanis, Sanjay Shakkottai, Wen-Sheng Chu:
Beyond First-Order Tweedie: Solving Inverse Problems using Latent Diffusion. 9472-9481
- You Wu, Kean Liu, Xiaoyue Mi, Fan Tang, Juan Cao, Jintao Li:
U-VAP: User-specified Visual Appearance Personalization via Decoupled Self Augmentation. 9482-9491
- Bingxin Ke, Anton Obukhov, Shengyu Huang, Nando Metzger, Rodrigo Caye Daudt, Konrad Schindler:
Repurposing Diffusion-Based Image Generators for Monocular Depth Estimation. 9492-9502
- Xin Kong, Shikun Liu, Xiaoyang Lyu, Marwan Taher, Xiaojuan Qi, Andrew J. Davison:
EscherNet: A Generative Model for Scalable View Synthesis. 9503-9513
- Khiem Vuong, N. Dinesh Reddy, Robert Tamburo, Srinivasa G. Narasimhan:
WALT3D: Generating Realistic Training Data from Time-Lapse Imagery for Reconstructing Dynamic Objects Under Occlusion. 9514-9524
- Yuanzhen Li, Fei Luo, Chunxia Xiao:
Diffusion-FOF: Single-View Clothed Human Reconstruction via Diffusion-Based Fourier Occupancy Field. 9525-9534
- Gwangbin Bae, Andrew J. Davison:
Rethinking Inductive Biases for Surface Normal Estimation. 9535-9545
- Mingqi Jiang, Saeed Khorram, Fuxin Li:
Comparing the Decision-Making Mechanisms by Transformers and CNNs via Explanation Methods. 9546-9555
- Xiang Yue, Yuansheng Ni, Tianyu Zheng, Kai Zhang, Ruoqi Liu, Ge Zhang, Samuel Stevens, Dongfu Jiang, Weiming Ren, Yuxuan Sun, Cong Wei, Botao Yu, Ruibin Yuan, Renliang Sun, Ming Yin, Boyuan Zheng, Zhenzhu Yang, Yibo Liu, Wenhao Huang, Huan Sun, Yu Su, Wenhui Chen:
MMMU: A Massive Multi-Discipline Multimodal Understanding and Reasoning Benchmark for Expert AGI. 9556-9567
- Shengbang Tong, Zhuang Liu, Yuexiang Zhai, Yi Ma, Yann LeCun, Saining Xie:
Eyes Wide Shut? Exploring the Visual Shortcomings of Multimodal LLMs. 9568-9578
- Xin Lai, Zhuotao Tian, Yukang Chen, Yanwei Li, Yuhui Yuan, Shu Liu, Jiaya Jia:
LISA: Reasoning Segmentation via Large Language Model. 9579-9589
- Yushi Hu, Otilia Stretcu, Chun-Ta Lu, Krishnamurthy Viswanathan, Kenji Hata, Enming Luo, Ranjay Krishna, Ariel Fuxman:
Visual Program Distillation: Distilling Tools and Programmatic Reasoning into Vision-Language Models. 9590-9601
- Bohan Yu, Jieji Ren, Jin Han, Feishi Wang, Jinxiu Liang, Boxin Shi:
EventPS: Real-Time Photometric Stereo Using an Event Camera. 9602-9611
- Xinyu Zhou, Peiqi Duan, Boyu Li, Chu Zhou, Chao Xu, Boxin Shi:
EvDiG: Event-guided Direct and Global Components Separation. 9612-9621
- Xiaolong Deng, Huisi Wu, Runhao Zeng, Jing Qin:
MemSAM: Taming Segment Anything Model for Echocardiography Video Segmentation. 9622-9631
- Guillaume Jaume, Lukas Oldenburg, Anurag Vaidya, Richard J. Chen, Drew F. K. Williamson, Thomas Peeters, Andrew H. Song, Faisal Mahmood:
Transcriptomics-Guided Slide Representation Learning in Computational Pathology. 9632-9644
- Mingyuan Meng , Dagan Feng, Lei Bi, Jinman Kim:
Correlation-aware Coarse-to-fine MLPs for Deformable Medical Image Registration. 9645-9654
- Pradyumna Reddy, Ismail Elezi, Jiankang Deng:
G3DR: Generative 3D Reconstruction in ImageNet. 9655-9665
- Haozhe Xie, Zhaoxi Chen, Fangzhou Hong, Ziwei Liu:
CityDreamer: Compositional Generative Model of Unbounded 3D Cities. 9666-9675
- Li Xu, Haoxuan Qu, Yujun Cai, Jun Liu:
6D-Diff: A Keypoint Diffusion Framework for 6D Object Pose Estimation. 9676-9686
- Lea Müller, Vickie Ye, Georgios Pavlakos, Michael J. Black, Angjoo Kanazawa:
Generative Proxemics: A Prior for 3D Social Interaction from Images. 9687-9697
- Hanzhe Hu, Zhizhuo Zhou, Varun Jampani, Shubham Tulsiani:
MVD-Fusion: Single-view 3D via Depth-consistent Multi-view Generation. 9698-9707
- Ziyao Zeng, Daniel Wang, Fengyu Yang, Hyoungseob Park, Stefano Soatto, Dong Lao, Alex Wong:
WorDepth: Variational Language Prior for Monocular Depth Estimation. 9708-9719
- Chuanxia Zheng, Andrea Vedaldi:
Free3D: Consistent Novel View Synthesis Without 3D Representation. 9720-9731
- Yu-Pei Song, Xiao Wu, Zhaoquan Yuanl, Jian-Jun Qiao, Qiang Peng:
PostureHMR: Posture Transformation for 3D Human Mesh Recovery. 9732-9741



- Linyi Jin, Nilesh Kulkarni, David F. Fouhey:
3DFIRES: Few Image 3D REconstruction for Scenes with Hidden Surfaces. 9742-9751
- Zizhang Li, Dor Litvak, Ruining Li, Yunzhi Zhang, Tomas Jakab, Christian Rupprecht, Shangzhe Wu, Andrea Vedaldi, Jiajun Wu:
Learning the 3D Fauna of the Web. 9752-9762
- Jie Tang, Fei-Peng Tian, Boshi An, Jian Li, Ping Tan:
Bilateral Propagation Network for Depth Completion. 9763-9772
- Heejoon Moon, Chunghwan Lee, Je Hyeong Hong:
Efficient Privacy-Preserving Visual Localization Using 3D Ray Clouds. 9773-9783
- Zehuan Huang, Hao Wen, Junting Dong, Yaohui Wang, Yangguang Li, Xinyuan Chen, Yan-Pei Cao, Ding Liang, Yu Qiao, Bo Dai, Lu Sheng:
EpiDiff: Enhancing Multi-View Synthesis via Localized Epipolar-Constrained Diffusion. 9784-9794
- Hmrishav Bandyopadhyay, Subhadeep Koley, Ayan Das, Ayan Kumar Bhunia, Aneeshan Sain, Pinaki Nath Chowdhury, Tao Xiang, Yi-Zhe Song:
Doodle Your 3D: from Abstract Freehand Sketches to Precise 3D Shapes. 9795-9805
- Linqing Zhao, Xiuwei Xu, Ziwei Wang, Yunpeng Zhang, Borui Zhang, Wenzhao Zheng, Dalong Du, Jie Zhou, Jiwen Lu:
LowRankOcc: Tensor Decomposition and Low-Rank Recovery for Vision-Based 3D Semantic Occupancy Prediction. 9806-9815
- Mohsen Yavartanoo, Sangmin Hong, Reyhaneh Neshatavar, Kyoung Mu Lee:
CNC-Net: Self-Supervised Learning for CNC Machining Operations. 9816-9825
- Georgios Pavlakos, Dandan Shan, Ilija Radosavovic, Angjoo Kanazawa, David Fouhey, Jitendra Malik:
Reconstructing Hands in 3D with Transformers. 9826-9836
- Keonhee Han, Dominik Muhle, Felix Wimbauer, Daniel Cremers:
Boosting Self-Supervision for Single-View Scene Completion via Knowledge Distillation. 9837-9847
- Rui Li, Tobias Fischer, Mattia Segù, Marc Pollefeys, Luc Van Gool, Federico Tombari:
Know Your Neighbors: Improving Single-View Reconstruction via Spatial Vision-Language Reasoning. 9848-9858
- Jin-Hwi Park, Chanhwi Jeong, Junoh Lee, Hae-Gon Jeon:
Depth Prompting for Sensor-Agnostic Depth Estimation. 9859-9869
- Xianghui Yang, Yan Zuo, Sameera Ramasinghe, Loris Bazzani, Gil Avraham, Anton van den Hengel:
ViewFusion: Towards Multi-View Consistency via Interpolated Denoising. 9870-9880
- Yizhi Wang, Wallace P. Lira, Wenqi Wang, Ali Mahdavi-Amiri, Hao Zhang:
Slice3D: Multi-Slice, Occlusion-Revealing, Single View 3D Reconstruction. 9881-9891
- Zike Wu, Pan Zhou, Xuanyu Yi, Xiaoding Yuan, Hanwang Zhang:
Consistent3D: Towards Consistent High-Fidelity Text-to-3D Generation with Deterministic Sampling Prior. 9892-9902
- Van Nguyen Nguyen, Thibault Groueix, Mathieu Salzmann, Vincent Lepetit:
GigaPose: Fast and Robust Novel Object Pose Estimation via One Correspondence. 9903-9913
- Lingteng Qiu, Guanying Chen, Xiaodong Gu, Qi Zuo, Mutian Xu, Yushuang Wu, Weihao Yuan, Zilong Dong, Liefeng Bo, Xiaoguang Han:
RichDreamer: A Generalizable Normal-Depth Diffusion Model for Detail Richness in Text-to-3D. 9914-9925
- Hao Ai, Lin Wang:
Elite360D: Towards Efficient 360 Depth Estimation via Semantic- and Distance-Aware Bi-Projection Fusion. 9926-9935
- Zechuan Zhang, Zongxin Yang, Yi Yang:
SIFU: Side-view Conditioned Implicit Function for Real-world Usable Clothed Human Reconstruction. 9936-9947
- Xuanyu Yi, Zike Wu, Qingshan Xu, Pan Zhou, Joo-Hwee Lim, Hanwang Zhang:
Diffusion Time-step Curriculum for One Image to 3D Generation. 9948-9958
- Yamei Chen, Yan Di, Guangyao Zhai, Fabian Manhardt, Chenyangguang Zhang, Ruida Zhang, Federico Tombari, Nassir Navab, Benjamin Busam:
SecondPose: SE(3)-Consistent Dual-Stream Feature Fusion for Category-Level Pose Estimation. 9959-9969
- Xiaoxiao Long, Yuan-Chen Guo, Cheng Lin, Yuan Liu, Zhiyang Dou, Lingjie Liu, Yuexin Ma, Song-Hai Zhang, Marc Habermann, Christian Theobalt, Wenping Wang:
Wonder3D: Single Image to 3D Using Cross-Domain Diffusion. 9970-9980
- Yifang Men, Biwen Lei, Yuan Yao, Miaomiao Cui, Zhouhui Lian, Xuansong Xie:
En3D: An Enhanced Generative Model for Sculpting 3D Humans from 2D Synthetic Data. 9981-9991
- Chenyangguang Zhang, Guanlong Jiao, Yan Di, Gu Wang, Ziqin Huang, Ruida Zhang, Fabian Manhardt, Bowen Fu, Federico Tombari, Xiangyang Ji:
MOHO: Learning Single-View Hand-Held Object Reconstruction with Multi-View Occlusion-Aware Supervision. 9992-10002
- Xianghui Xie, Bharat Lal Bhatnagar, Jan Eric Lenssen, Gerard Pons-Moll:
Template Free Reconstruction of Human-object Interaction with Procedural Interaction Generation. 10003-10015



- Zhenyu Li, Shariq Farooq Bhat, Peter Wonka:
PatchFusion: An End-to-End Tile-Based Framework for High-Resolution Monocular Metric Depth Estimation. 10016-10025
- Yash Kant, Aliaksandr Siarohin, Ziyi Wu, Michael Vasilkovsky, Guocheng Qian, Jian Ren, Riza Alp Güler, Bernard Ghanem, Sergey Tulyakov, Igor Gilitschenski:
SPAD: Spatially Aware Multi-View Diffusers. 10026-10038
- Sungphil Moon, Hyeontae Son, Dongcheol Hur, Sangwook Kim:
GenFlow: Generalizable Recurrent Flow for 6D Pose Refinement of Novel Objects. 10039-10049
- Zixuan Huang, Justin Johnson, Shoubhik Debnath, James M. Rehg, Chao-Yuan Wu:
PointInfinity: Resolution-Invariant Point Diffusion Models. 10050-10060
- Zixuan Huang, Stefan Stojanov, Anh Thai, Varun Jampani, James M. Rehg:
ZeroShape: Regression-Based Zero-Shot Shape Reconstruction. 10061-10071
- Minghua Liu, Ruoxi Shi, Linghao Chen, Zhuoyang Zhang, Chao Xu, Xinyue Wei, Hansheng Chen, Chong Zeng, Jiayuan Gu, Hao Su:
One-2-3-45++: Fast Single Image to 3D Objects with Consistent Multi-View Generation and 3D Diffusion. 10072-10083
- Xiangjun Gao, Xiaoyu Li, Chaopeng Zhang, Qi Zhang, Yanpei Cao, Ying Shan, Long Quan:
ConTex-Human: Free-View Rendering of Human from a Single Image with Texture-Consistent Synthesis. 10084-10094
- Junwen Huang, Hao Yu, Kuan-Ting Yu, Nassir Navab, Slobodan Ilic, Benjamin Busam:
MatchU: Matching Unseen Objects for 6D Pose Estimation from RGB-D Images. 10095-10105
- Luigi Piccinelli, Yung-Hsu Yang, Christos Sakaridis, Mattia Segù, Siyuan Li, Luc Van Gool, Fisher Yu:
UniDepth: Universal Monocular Metric Depth Estimation. 10106-10116
- Zixiong Huang, Qi Chen, Libo Sun, Yifan Yang, Naizhou Wang, Qi Wu, Mingkui Tan:
G-NeRF: Geometry-enhanced Novel View Synthesis from Single-View Images. 10117-10126
- Yifang Men, Hanxi Liu, Yuan Yao, Miaomiao Cui, Xuansong Xie, Zhouhui Lian:
3DToonify: Creating Your High-Fidelity 3D Stylized Avatar Easily from 2D Portrait Images. 10127-10137
- Junda Cheng, Wei Yin, Kaixuan Wang, Xiaozhi Chen, Shijie Wang, Xin Yang:
Adaptive Fusion of Single-View and Multi-View Depth for Autonomous Driving. 10138-10147
- Yongliang Lin, Yongzhi Su, Praveen Nathan, Sandeep Inuganti, Yan Di, Martin Sundermeyer, Fabian Manhardt, Didier Stricker, Jason R. Rambach, Yu Zhang:
HiPose: Hierarchical Binary Surface Encoding and Correspondence Pruning for RGB-D 6DoF Object Pose Estimation. 10148-10158
- Hao Xu, Haipeng Li, Yinqiao Wang, Shuaicheng Liu, Chi-Wing Fu:
HandBooster: Boosting 3D Hand-Mesh Reconstruction by Conditional Synthesis and Sampling of Hand-Object Interactions. 10159-10169
- Songchun Zhang, Yibo Zhang, Quan Zheng, Rui Ma, Wei Hua, Hujun Bao, Weiwei Xu, Changqing Zou:
3D-SceneDreamer: Text-Driven 3D-Consistent Scene Generation. 10170-10180
- Wonbong Jang, Lourdes Agapito:
NViST: In the Wild New View Synthesis from a Single Image with Transformers. 10181-10193
- Ziyu Wan, Despoina Paschalidou, Ian Huang, Hongyu Liu, Bokui Shen, Xiaoyu Xiang, Jing Liao, Leonidas J. Guibas:
CAD : Photorealistic 3D Generation via Adversarial Distillation. 10194-10207
- Stanislaw Szymanowicz, Christian Rupprecht, Andrea Vedaldi:
Splatter Image: Ultra-Fast Single-View 3D Reconstruction. 10208-10217
- Hyeongjin Nam, Daniel Sungho Jung, Gyeongsik Moon, Kyoung Mu Lee:
Joint Reconstruction of 3D Human and Object via Contact-Based Refinement Transformer. 10218-10227
- Cheng Chen, Xiaofeng Yang, Fan Yang, Chengzeng Feng, Zhoujie Fu, Chuan-Sheng Foo, Guosheng Lin, Fayao Liu:
Sculpt3D: Multi-View Consistent Text-to-3D Generation with Sparse 3D Prior. 10228-10237
- Tianfu Wang, Guosheng Hu, Hongguang Wang:
Object Pose Estimation via the Aggregation of Diffusion Features. 10238-10247
- Longfei Yan, Pei Yan, Shengzhou Xiong, Xuanyu Xiang, Yihua Tan:
MonoCD: Monocular 3D Object Detection with Complementary Depths. 10248-10257
- Norman Müller, Katja Schwarz, Barbara Rössle, Lorenzo Porzi, Samuel Rota Bulò, Matthias Nießner, Peter Kortscheder:
MultiDiff: Consistent Novel View Synthesis from a Single Image. 10258-10268
- Abhinav Kumar, Yuliang Guo, Xinyu Huang, Liu Ren, Xiaoming Liu:
SeaBird: Segmentation in Bird's View with Dice Loss Improves Monocular 3D Detection of Large Objects. 10269-10280
- Liang Peng, Junkai Xu, Haoran Cheng, Zheng Yang, Xiaopei Wu, Wei Qian, Wenxiao Wang, Boxi Wu, Deng Cai:
Learning Occupancy for Monocular 3D Object Detection. 10281-10292
- Zhenggang Tang, Zhongzheng Ren, Xiaoming Zhao, Bowen Wen, Jonathan Tremblay, Stan Birchfield, Alexander G. Schwing:
NeRFDeformer: NeRF Transformation from a Single View via 3D Scene Flows. 10293-10303



- Kennard Yanting Chan, Fayao Liu, Guosheng Lin, Chuan Sheng Foo, Weisi Lin:
R-Cyclic Diffuser: Reductive and Cyclic Latent Diffusion for 3D Clothed Human Digitalization. 10304-10313
- Fengyun Wang, Qianru Sun, Dong Zhang, Jinhui Tang:
Unleashing Network Potentials for Semantic Scene Completion. 10314-10323
- Zi-Xin Zou, Zhipeng Yu, Yuan-Chen Guo, Yangguang Li, Ding Liang, Yan-Pei Cao, Song-Hai Zhang:
Triplane Meets Gaussian Splatting: Fast and Generalizable Single-View 3D Reconstruction with Transformers. 10324-10335
- Phong Tran, Egor Zakharov, Long-Nhat Ho, Anh Tuan Tran, Liwen Hu, Hao Li:
VOODOO 3D: Volumetric Portrait Disentanglement for One-Shot 3D Head Reenactment. 10336-10348
- Simon Niedermayr, Josef Stumpfegger, Rüdiger Westermann:
Compressed 3D Gaussian Splatting for Accelerated Novel View Synthesis. 10349-10358
- Xiyi Chen, Marko Mihajlovic, Shaofei Wang, Sergey Prokudin, Siyu Tang:
Morphable Diffusion: 3D-Consistent Diffusion for Single-image Avatar Creation. 10359-10370
- Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, Hengshuang Zhao:
Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data. 10371-10381
- Mehmet Aygün, Oisin Mac Aodha:
SAOR: Single-View Articulated Object Reconstruction. 10382-10391
- Haozhe Qi, Chen Zhao, Mathieu Salzmann, Alexander Mathis:
HOISDF: Constraining 3D Hand-Object Pose Estimation with Global Signed Distance Fields. 10392-10402
- Jihyun Kim, Changjae Oh, Hoseok Do, Soohyun Kim, Kwanghoon Sohn:
Diffusion-Driven GAN Inversion for Multi-Modal Face Image Generation. 10403-10412
- Juan Luis Gonzalez Bello, Munchurl Kim:
Novel View Synthesis with View-Dependent Effects from a Single Image. 10413-10423
- Xingqun Qi, Jiahao Pan, Peng Li, Ruibin Yuan, Xiaowei Chi, Mengfei Li, Wenhan Luo, Wei Xue, Shanghang Zhang, Qifeng Liu, Yike Guo:
Weakly-Supervised Emotion Transition Learning for Diverse 3D Co-Speech Gesture Generation. 10424-10434
- Cheng Sun, Wei-En Tai, Yu-Lin Shih, Kuan-Wei Chen, Yong-Jing Syu, Kent Selwyn The, Yu-Chiang Frank Wang, Hwann-Tzong Chen:
Seg2Reg: Differentiable 2D Segmentation to 1D Regression Rendering for 360 Room Layout Reconstruction. 10435-10445
- Hoang Chuong Nguyen, Tianyu Wang, José M. Álvarez, Miaomiao Liu:
Mining Supervision for Dynamic Regions in Self-Supervised Monocular Depth Estimation. 10446-10455
- Yuming Gu, Hongyi Xu, You Xie, Guoxian Song, Yichun Shi, Di Chang, Jing Yang, Linjie Luo:
DiffPortrait3D: Controllable Diffusion for Zero-Shot Portrait View Synthesis. 10456-10465
- Mosam Dabhi, László A. Jeni, Simon Lucey:
3D-LFM: Lifting Foundation Model. 10466-10475
- Yuelong Li, Yafei Mao, Raja Bala, Sunil Hadap:
MRC-Net: 6-DoF Pose Estimation with MultiScale Residual Correlation. 10476-10486
- Biwen Lei, Kai Yu, Mengyang Feng, Miaomiao Cui, Xuansong Xie:
DiffusionGAN3D: Boosting Text-guided 3D Generation and Domain Adaptation by Combining 3D GANs and Diffusion Priors. 10487-10497
- Leyuan Liu, Yuhua Li, Yunqi Gao, Changxin Gao, Yuanyuan Liu, Jingying Chen:
VS: Reconstructing Clothed 3D Human from Single Image via Vertex Shift. 10498-10507
- Xueying Jiang, Sheng Jin, Lewei Lu, Xiaoqin Zhang, Shijian Lu:
Weakly Supervised Monocular 3D Detection with a Single-View Image. 10508-10518
- Jaeho Moon, Juan Luis Gonzalez Bello, Byeongjun Kwon, Munchurl Kim:
From-Ground-To-Objects: Coarse-to-Fine Self-supervised Monocular Depth Estimation of Dynamic Objects with Ground Contact Prior. 10519-10529
- Andrea Ramazzina, Stefanie Walz, Pragyan Dahal, Mario Bijelic, Felix Heide:
Gated Fields: Learning Scene Reconstruction from Gated Videos. 10530-10541
- Yunhao Li, Xiaodong Wang, Ping Wang, Xin Yuan, Peidong Liu:
SCINeRF: Neural Radiance Fields from a Snapshot Compressive Image. 10542-10552
- Mi-Gyeong Gwon, Gi-Mun Um, Won-Sik Cheong, Wonjun Kim:
Instance-Aware Contrastive Learning for Occluded Human Mesh Reconstruction. 10553-10562
- Minghao Yin, Shangzhe Wu, Kai Han:
IBD-SLAM: Learning Image-Based Depth Fusion for Generalizable SLAM. 10563-10573
- Sangmin Woo, Byeongjun Park, Hyojun Go, Jin-Young Kim, Changick Kim:
HarmonyView: Harmonizing Consistency and Diversity in One-Image-to-3D. 10574-10584
- Hong Li, Yutang Feng, Song Xue, Xuhui Liu, Bohan Zeng, Shanglin Li, Boyu Liu, Jianzhuang Liu, Shumin Han, Baochang Zhang:
UV-IDM: Identity-Conditioned Latent Diffusion Model for Face UV-Texture Generation. 10585-10595



- Fan Yang, Tianyi Chen, Xiaosheng He, Zhongang Cai, Lei Yang, Si Wu, Guosheng Lin:
AttriHuman-3D: Editable 3D Human Avatar Generation with Attribute Decomposition and Indexing. 10596-10605
- Lior Talker, Aviad Cohen, Erez Yosef, Alexandra Dana, Michael Dinerstein:
Mind The Edge: Refining Depth Edges in Sparsely-Supervised Monocular Depth Estimation. 10606-10616
- Chenfeng Xu, Huan Ling, Sanja Fidler, Or Litany:
3DiffTection: 3D Object Detection with Geometry-Aware Diffusion Features. 10617-10627
- Haiyang Xu, Yu Lei, Zeyuan Chen, Xiang Zhang, Yue Zhao, Yilin Wang, Zhuowen Tu:
Bayesian Diffusion Models for 3D Shape Reconstruction. 10628-10638
- Maximilian Pittner, Joel Janai, Alexandru Paul Condurache:
LaneCPP: Continuous 3D Lane Detection Using Physical Priors. 10639-10648
- Seungwook Kim, Kejie Li, Xueqing Deng, Yichun Shi, Minsu Cho, Peng Wang:
Enhancing 3D Fidelity of Text-to-3D using Cross-View Correspondences. 10649-10658
- Yasiru Ranasinghe, Deepti Hegde, Vishal M. Patel:
MonoDiff: Monocular 3D Object Detection and Pose Estimation with Diffusion Models. 10659-10670
- Yifan Yang, Dong Liu, Shuhai Zhang, Zeshuai Deng, Zixiong Huang, Mingkui Tan:
HiLo: Detailed and Robust 3D Clothed Human Reconstruction with High-and Low-Frequency Information of Parametric Models. 10671-10681
- Jimin Xu, Tianbao Wang, Tao Jin, Shengyu Zhang, Dongjie Fu, Zhe Wang, Jiangjing Lyu, Chengfei Lv, Chaoyue Niu, Zhou Yu, Zhou Zhao, Fei Wu:
MPOD123: One Image to 3D Content Generation Using Mask-Enhanced Progressive Outline-to-Detail Optimization. 10682-10692
- Linfang Zheng, Tze Ho Elden Tse, Chen Wang, Yinghan Sun, Hua Chen, Ales Leonardis, Wei Zhang, Hyung Jin Chang:
GeoReF: Geometric Alignment Across Shape Variation for Category-level Object Pose Refinement. 10693-10703
- Weikang Wang, Dongliang Cao, Florian Bernard:
Unsupervised 3D Structure Inference from Category-Specific Image Collections. 10704-10714
- Devikalyan Das, Christopher Wewer, Raza Yunus, Eddy Ilg, Jan Eric Lenssen:
Neural Parametric Gaussians for Monocular Non-Rigid Object Reconstruction. 10715-10725
- Minje Kim, Tae-Kyun Kim:
BiTT: Bi-Directional Texture Reconstruction of Interacting Two Hands from a Single Image. 10726-10735
- Yunxiao Shi, Manish Kumar Singh, Hong Cai, Fatih Porikli:
DeCoTR: Enhancing Depth Completion with 2D and 3D Attentions. 10736-10746
- Simon Giebenhain, Tobias Kirschstein, Markos Georgopoulos, Martin Rünz, Lourdes Agapito, Matthias Nießner:
MonoNPHM: Dynamic Head Reconstruction from Monocular Videos. 10747-10758
- George Cazenavette, Avneesh Sud, Thomas Leung, Ben Usman:
FakeInversion: Learning to Detect Images from Unseen Text-to-Image Models by Inverting Stable Diffusion. 10759-10769
- Huan Liu, Zichang Tan, Chuangchuang Tan, Yunchao Wei, Jingdong Wang, Yao Zhao:
Forgery-aware Adaptive Transformer for Generalizable Synthetic Image Detection. 10770-10780
- Chenfan Qu, Yiwu Zhong, Chongyu Liu, Guitao Xu, Dezhi Peng, Fengjun Guo, Lianwen Jin:
Towards Modern Image Manipulation Localization: A Large-Scale Dataset and Novel Methods. 10781-10790
- Vishal Asnani, John P. Collomosse, Tu Bui, Xiaoming Liu, Shruti Agarwal:
ProMark: Proactive Diffusion Watermarking for Causal Attribution. 10802-10811
- Xiaoyu Wu, Yang Hua, Chumeng Liang, Jiaru Zhang, Hao Wang, Tao Song, Haibing Guan:
CGI-DM: Digital Copyright Authentication for Diffusion Models via Contrasting Gradient Inversion. 10812-10821
- Zhixuan Liu, Peter Schaldenbrand, Beverley-Claire Okogwu, Wenxuan Peng, Wenxuan Peng, Youngsik Yun, Andrew Hundt, Jihie Kim, Jean Oh:
SCoFT: Self-Contrastive Fine-Tuning for Equitable Image Generation. 10822-10832
- Tianwei Chen, Yusuke Hirota, Mayu Otani, Noa Garcia, Yuta Nakashima:
Would Deep Generative Models Amplify Bias in Future Models? 10833-10843
- Zichen Miao, Jiang Wang, Ze Wang, Zhengyuan Yang, Lijuan Wang, Qiang Qiu, Zicheng Liu:
Training Diffusion Models Towards Diverse Image Generation with Reinforcement Learning. 10844-10853
- Zaid Khan, Yun Fu:
Consistency and Uncertainty: Identifying Unreliable Responses From Black-Box Vision-Language Models for Selective Visual Question Answering. 10854-10863
- Julie Tores, Lucile Sassatelli, Hui-Yin Wu, Clement Bergman, Lea Andolfi, Victor Ecrement, Frédéric Precioso, Thierry Devars, Magali Guaresi, Virginie Julliard, Sarah Lecossais:
Visual Objectification in Films: Towards a New AI Task for Video Interpretation. 10864-10874
- Kartik Thakral, Shashikant Prasad, Stuti Aswani, Mayank Vatsa, Richa Singh:
ToonerGAN: Reinforcing GANs for Obfuscating Automated Facial Indexing. 10875-10884
- Bor-Shiun Wang, Chien-Yi Wang, Wei-Chen Chiu:
MCPNet: An Interpretable Classifier via Multi-Level Concept Prototypes. 10885-10894



- Matthew Kowal, Richard P. Wildes, Konstantinos G. Derpanis:
Visual Concept Connectome (VCC): Open World Concept Discovery and Their Interlayer Connections in Deep Models. 10895-10905
- Zeliang Zhang, Mingqian Feng, Zhiheng Li, Chenliang Xu:
Discover and Mitigate Multiple Biased Subgroups in Image Classifiers. 10906-10915
- Keke Tang, Chao Hou, Weilong Peng, Runnan Chen, Peican Zhu, Wenping Wang, Zhihong Tian:
CORES: Convolutional Response-based Score for Out-of-distribution Detection. 10916-10925
- Junyi Wu, Bin Duan, Weitai Kang, Hao Tang, Yan Yan:
Token Transformation Matters: Towards Faithful Post-Hoc Explanation for Vision Transformer. 10926-10935
- Junyi Wu, Weitai Kang, Hao Tang, Yuan Hong, Yan Yan:
On the Faithfulness of Vision Transformer Explanations. 10936-10945
- Matthew Kowal, Achal Dave, Rares Ambrus, Adrien Gaidon, Konstantinos G. Derpanis, Pavel Tokmakov:
Understanding Video Transformers via Universal Concept Discovery. 10946-10956
- Namitha Padmanabhan, Matthew Gwilliam, Pulkit Kumar, Shishira R. Maiya, Max Ehrlich, Abhinav Shrivastava:
Explaining the Implicit Neural Canvas: Connecting Pixels to Neurons by Tracing Their Contributions. 10957-10967
- Yong Hyun Ahn, Hyeon Bae Kim, Seong Tae Kim:
WWW: A Unified Framework for Explaining what, Where and why of Neural Networks by Interpretation of Neuron Concepts. 10968-10977
- Hae Jin Song, Mahyar Khayatkhoei, Wael AbdAlmageed:
ManiFPT: Defining and Analyzing Fingerprints of Generative Models. 10971-10981
- Prathyush Poduval, Zhuowen Zou, Mohsen Imani:
HDQMF: Holographic Feature Decomposition using Quantum Algorithms. 10978-10987
- Revoti Prasad Bora, Philipp Terhörst, Raymond N. J. Veldhuis, Raghavendra Ramachandra, Kiran B. Raja:
SLICE: Stabilized LIME for Consistent Explanations for Image Classification. 10988-10996
- Hmrishav Bandyopadhyay, Pinaki Nath Chowdhury, Ayan Kumar Bhunia, Aneeshan Sain, Tao Xiang, Yi-Zhe Song:
What Sketch Explainability Really Means for Downstream Tasks? 10997-11008
- Shizhan Gong, Qi Dou, Farzan Farnia:
Structured Gradient-Based Interpretations via Norm-Regularized Adversarial Training. 11009-11018
- Ping Chen, Xingpeng Zhang, Chengtao Zhou, Dichao Fan, Peng Tu, Le Zhang, Yanlin Qian:
Learning Triangular Distribution in Visual World. 11019-11029
- Chenming Shang, Shiji Zhou, Hengyuan Zhang, Xinzhe Ni, Yujiu Yang, Yuwang Wang:
Incremental Residual Concept Bottleneck Models. 11030-11040
- Omer Yair, Elias Nehme, Tomer Michaeli:
Uncertainty Visualization via Low-Dimensional Posterior Projections. 11041-11051
- Hanjing Wang, Qiang Ji:
Epistemic Uncertainty Quantification for Pretrained Neural Networks. 11052-11061
- Alessandro Achille, Greg Ver Steeg, Tian Yu Liu, Matthew Trager, Carson Klingenberg, Stefano Soatto:
Interpretable Measures of Conceptual Similarity by Complexity-Constrained Descriptive Auto-Encoding. 11062-11071
- Townim Faisal Chowdhury, Kewen Liao, Vu Minh Hieu Phan, Minh-Son To, Yutong Xie, Kevin Hung, David Ross, Anton van den Hengel, Johan W. Verjans, Zhibin Liao:
CAPE: CAM as a Probabilistic Ensemble for Enhanced DNN Interpretation. 11072-11081
- Younghyun Kim, Sangwoo Mo, Minkyu Kim, Kyungmin Lee, Jaeho Lee, Jinwoo Shin:
Discovering and Mitigating Visual Biases Through Keyword Explanation. 11082-11092
- Maximilian Augustin, Yannic Neuhaus, Matthias Hein:
DiG-IN: Diffusion Guidance for Investigating Networks - Uncovering Classifier Differences, Neuron Visualisations, and Visual Counterfactual Explanations. 11093-11103
- Xiaoyu Liu, Miaomiao Cai, Yinda Chen, Yueyi Zhang, Te Shi, Ruobing Zhang, Xuejin Chen, Zhiwei Xiong:
Cross-dimension Affinity Distillation for 3D EM Neuron Segmentation. 11104-11113
- Yiwen Ye, Yutong Xie, Jianpeng Zhang, Ziyang Chen, Qi Wu, Yong Xia:
Continual Self-Supervised Learning: Towards Universal Multi-Modal Medical Data Representation Learning. 11114-11124
- Yuelin Zhang, Pengyu Zheng, Wanquan Yan, Chengyu Fang, Shing Shin Cheng:
A Unified Framework for Microscopy Defocus Deblur with Multi-Pyramid Transformer and Contrastive Learning. 11125-11136
- Haoran Lai, Qingsong Yao, Zihang Jiang, Rongsheng Wang, Zhiyang He, Xiaodong Tao, S. Kevin Zhou:
CARZero: Cross-Attention Alignment for Radiology Zero-Shot Classification. 11137-11146
- Qi Chen, Xiaoxi Chen, Haorui Song, Zhiwei Xiong, Alan L. Yuille, Chen Wei, Zongwei Zhou:
Towards Generalizable Tumor Synthesis. 11147-11158
- Marianne Rakic, Hallee E. Wong, Jose Javier Gonzalez Ortiz, Beth A. Cimini, John V. Guttag, Adrian V. Dalca:
Tyche: Stochastic in-Context Learning for Medical Image Segmentation. 11159-11173



- Yuanhao Cai, Jiahao Wang, Alan L. Yuille, Zongwei Zhou, Angtian Wang:
Structure-Aware Sparse-View X-Ray 3D Reconstruction. 11174-11183
- Ziyang Chen, Yongsheng Pan, Yiwen Ye, Mengkang Lu, Yong Xia:
Each Test Image Deserves A Specific Prompt: Continual Test-Time Adaptation for 2D Medical Image Segmentation. 11184-11193
- Yunhe Gao:
Training Like a Medical Resident: Context-Prior Learning Toward Universal Medical Image Segmentation. 11194-11204
- Yiqun Lin, Jiewen Yang, Hualiang Wang, Xinpeng Ding, Wei Zhao, Xiaomeng Li:
C²RV: Cross-Regional and Cross-View Learning for Sparse-View CBCT Reconstruction. 11205-11214
- Tony C. W. Mok, Zi Li, Yunhao Bai, Jianpeng Zhang, Wei Liu, Yan-Jie Zhou, Ke Yan, Dakai Jin, Yu Shi, Xiaoli Yin, Le Lu, Ling Zhang:
Modality-Agnostic Structural Image Representation Learning for Deformable Multi-Modality Medical Image Registration. 11215-11225
- Saarthak Kapse, Pushpak Pati, Srijan Das, Jingwei Zhang, Chao Chen, Maria Vakalopoulou, Joel H. Saltz, Dimitris Samaras, Rajarsi R. Gupta, Prateek Prasanna:
SI-MIL: Taming Deep MIL for Self-Interpretability in Gigapixel Histopathology. 11226-11237
- Weiwei Cao, Jianpeng Zhang, Yingda Xia, Tony C. W. Mok, Zi Li, Xianghua Ye, Le Lu, Jian Zheng, Yuxing Tang, Ling Zhang:
Bootstrapping Chest CT Image Understanding by Distilling Knowledge from X-Ray Expert Models. 11238-11247
- Jiangbo Shi, Chen Li, Tieliang Gong, Yefeng Zheng, Huazhu Fu:
ViLa-MIL: Dual-scale Vision-Language Multiple Instance Learning for Whole Slide Image Classification. 11248-11258
- Jiahua Li, Jiuyang Dong, Shenjin Huang, Xi Li, Junjun Jiang, Xiaopeng Fan, Yongbing Zhang:
Virtual Immunohistochemistry Staining for Histological Images Assisted by Weakly-supervised Learning. 11259-11268
- Mohammad Reza Hosseinzadeh Taher, Michael B. Gotway, Jianming Liang:
Representing Part-Whole Hierarchies in Foundation Models by Learning Localizability, Composability, and Decomposability from Anatomy via Self-Supervision. 11269-11281
- Chong Yin, Siqi Liu, Fei Lyu, Jiahao Lu, Sune Darkner, Vincent Wai-Sun Wong, Pong C. Yuen:
XFibrosis: Explicit Vessel-Fiber Modeling for Fibrosis Staging from Liver Pathology Images. 11282-11291
- Chong Yin, Siqi Liu, Kaiyang Zhou, Vincent Wai-Sun Wong, Pong C. Yuen:
Prompting Vision Foundation Models for Pathology Image Analysis. 11292-11301
- Junde Wu, Min Xu:
One-Prompt to Segment All Medical Images. 11302-11312
- Jiateng Shou, Zeyu Xiao, Shiyu Deng, Wei Huang, Peiyao Shi, Ruobing Zhang, Zhiwei Xiong, Feng Wu:
Learning Large-Factor EM Image Super-Resolution with Generative Priors. 11313-11322
- Jiawen Li, Yuxuan Chen, Hongbo Chu, Qiehe Sun, Tian Guan, Anjia Han, Yonghong He:
Dynamic Graph Representation with Knowledge-Aware Attention for Histopathology Whole Slide Image Analysis. 11323-11332
- Shizun Wang, Songhua Liu, Zhenxiong Tan, Xincho Wang:
MindBridge: A Cross-Subject Brain Decoding Framework. 11333-11342
- Wenhao Tang, Fengtao Zhou, Sheng Huang, Xiang Zhu, Yi Zhang, Bo Liu:
Feature Re-Embedding: Towards Foundation Model-Level Performance in Computational Pathology. 11343-11352
- JungEun Kim, Hangyul Yoon, Geondo Park, Kyungsu Kim, Eunho Yang:
Data-Efficient Unsupervised Interpolation Without Any Intermediate Frame for 4D Medical Images. 11353-11364
- Guangyuan Li, Chen Rao, Juncheng Mo, Zhanjie Zhang, Wei Xing, Lei Zhao:
Rethinking Diffusion Model for Multi-Contrast MRI Super-Resolution. 11365-11374
- Chaoqin Huang, Aofan Jiang, Jinghao Feng, Ya Zhang, Xincho Wang, Yanfeng Wang:
Adapting Visual-Language Models for Generalizable Anomaly Detection in Medical Images. 11375-11385
- Yankai Jiang, Zhongzhen Huang, Rongzhao Zhang, Xiaofan Zhang, Shaoting Zhang:
ZePT: Zero-Shot Pan-Tumor Segmentation via Query-Disentangling and Self-Prompting. 11386-11397
- Hao Li, Ying Chen, Yifei Chen, Rongshan Yu, Wenxian Yang, Liansheng Wang, Bowen Ding, Yuchen Han:
Generalizable Whole Slide Image Classification with Fine-Grained Visual-Semantic Interaction. 11398-11407
- Huyong Wang, Huisi Wu, Jing Qin:
Incremental Nuclei Segmentation from Histopathological Images via Future-class Awareness and Compatibility-inspired Distillation. 11408-11417
- Siyao Jiang, Huisi Wu, Junyang Chen, Qin Zhang, Jing Qin:
PH-Net: Semi-Supervised Breast Lesion Segmentation via Patch-Wise Hardness. 11418-11427
- Marius Schmidt-Mengin, Alexis Benichoux, Shibeshih Mitiku Belachew, Nikos Komodakis, Nikos Paragios:
ToNNO: Tomographic Reconstruction of a Neural Network's Output for Weakly Supervised Segmentation of 3D Medical Images. 11428-11438



- Jiayi Chen, Benteng Ma, Hengfei Cui, Yong Xia:
Think Twice Before Selection: Federated Evidential Active Learning for Medical Image Analysis with Domain Shifts. 11439-11449
- Sajid Javed, Arif Mahmood, Iyyakutti Iyappan Ganapathi, Fayaz Ali Dharejo, Naoufel Werghi, Mohammed Bennamoun:
CPLIP: Zero-Shot Learning for Histopathology with Comprehensive Vision-Language Alignment. 11450-11459
- Mude Hui, Zihao Wei, Hongru Zhu, Fei Xia, Yuyin Zhou:
MicroDiffusion: Implicit Representation-Guided Diffusion for 3D Reconstruction from Limited 2D Microscopy Projections. 11460-11469
- Yicheng Wu, Xiangde Luo, Zhe Xu, Xiaoqing Guo, Lie Ju, Zongyuan Ge, Wenjun Liao, Jianfei Cai:
Diversified and Personalized Multi-Rater Medical Image Segmentation. 11470-11479
- Ju-Hyeon Nam, Nur Suriza Syazwany, Su Jung Kim, Sang-Chul Lee:
Modality-Agnostic Domain Generalizable Medical Image Segmentation by Multi-Frequency in Multi-Scale Attention. 11480-11491
- Vu Minh Hieu Phan, Yutong Xie, Yuankai Qi, Lingqiao Liu, Liyang Liu, Bowen Zhang, Zhibin Liao, Qi Wu, Minh-Son To, Johan W. Verjans:
Decomposing Disease Descriptions for Enhanced Pathology Detection: A Multi-Aspect Vision-Language Pre-Training Framework. 11492-11501
- Chenlu Zhan, Yu Lin, Gaoang Wang, Hongwei Wang, Jian Wu:
MedM2G: Unifying Medical Multi-Modal Generation via Cross-Guided Diffusion with Visual Invariant. 11502-11512
- Morteza Ghahremani, Mohammad Khateri, Bailiang Jian, Benedikt Wiestler, Ehsan Adeli, Christian Wachinger:
H-ViT: A Hierarchical Vision Transformer for Deformable Image Registration. 11513-11523
- Jianan Fan, Dongnan Liu, Hang Chang, Heng Huang, Mei Chen, Weidong Cai:
Seeing Unseen: Discover Novel Biomedical Concepts via Geometry-Constrained Probabilistic Modeling. 11524-11534
- Sean I. Young, Yaël Balbastre, Bruce Fischl, Polina Golland, Juan Eugenio Iglesias:
Fully Convolutional Slice-to-Volume Reconstruction for Single-Stack MRI. 11535-11545
- Tai Ma, Suwei Zhang, Jiafeng Li, Ying Wen:
IIRP-Net: Iterative Inference Residual Pyramid Network for Enhanced Image Registration. 11546-11555
- Nicolas Bourriez, Ihab Bendidi, Ethan Cohen, Gabriel Watkinson, Maxime Sanchez, Guillaume Bollot, Auguste Genovesio:
ChAda-ViT : Channel Adaptive Attention for Joint Representation Learning of Heterogeneous Microscopy Image. 11556-11565
- Andrew H. Song, Richard J. Chen, Tong Ding, Drew F. K. Williamson, Guillaume Jaume, Faisal Mahmood:
Morphological Prototyping for Unsupervised Slide Representation Learning in Computational Pathology. 11566-11578
- Guillaume Jaume, Anurag Vaidya, Richard J. Chen, Drew F. K. Williamson, Paul Pu Liang, Faisal Mahmood:
Modeling Dense Multimodal Interactions Between Biological Pathways and Histology for Survival Prediction. 11579-11590
- Youngmin Chung, Ji Hun Ha, Kyeong Chan Im, Joo Sang Lee:
Accurate Spatial Gene Expression Prediction by Integrating Multi-Resolution Features. 11591-11600
- Bo Zou, Shaofeng Wang, Hao Liu, Gaoyue Sun, Yajie Wang, Feifei Zuo, Chengbin Quan, Youjian Zhao:
Teeth-SEG: An Efficient Instance Segmentation Framework for Orthodontic Treatment Based on Multi-Scale Aggregation and Anthropic Prior Knowledge. 11601-11610
- Yuhang Zhou, Haolin Li, Siyuan Du, Jiangchao Yao, Ya Zhang, Yanfeng Wang:
Low-Rank Knowledge Decomposition for Medical Foundation Models. 11611-11620
- Bin Pu, Liwen Wang, Jiewen Yang, Guannan He, Xingbo Dong, Shengli Li, Ying Tan, Ming Chen, Zhe Jin, Kenli Li, Xiaomeng Li:
M³-UDA: A New Benchmark for Unsupervised Domain Adaptive Fetal Cardiac Structure Detection. 11621-11630
- Wei Fang, Yuxing Tang, Heng Guo, Mingze Yuan, Tony C. W. Mok, Ke Yan, Jiawen Yao, Xin Chen, Zaiyi Liu, Le Lu, Ling Zhang, Minfeng Xu:
CycleINR: Cycle Implicit Neural Representation for Arbitrary-Scale Volumetric Super-Resolution of Medical Data. 11631-11641
- Qinghe Ma, Jian Zhang, Lei Qi, Qian Yu, Yinghuan Shi, Yang Gao:
Constructing and Exploring Intermediate Domains in Mixed Domain Semi-supervised Medical Image Segmentation. 11642-11651
- Yutong Xie, Qi Chen, Sinuo Wang, Minh-Son To, Iris Lee, Ee Win Khoo, Kerolos Hendy, Daniel Koh, Yong Xia, Qi Wu:
PairAug: What Can Augmented Image-Text Pairs Do for Radiology? 11652-11661
- Vivek Gopalakrishnan, Neel Dey, Polina Golland:
Intraoperative 2D/3D Image Registration via Differentiable X-Ray Rendering. 11662-11672
- Jun Wang:
Mudslide: A Universal Nuclear Instance Segmentation Method. 11673-11682
- Saghir Alfasly, Abubakr Shafique, Peyman Nejat, Jibran A. Khan, Areej Alsaafin, Ghazal Alabtah, Hamid R. Tizhoosh:
Rotation-Agnostic Image Representation Learning for Digital Pathology. 11683-11693



- Wei Shao, Yangyang Shi, Daoqiang Zhang, Junjie Zhou, Peng Wan:
Tumor Micro-Environment Interactions Guided Graph Learning for Survival Analysis of Human Cancers from Whole-Slide Pathological Images. 11694-11703
- Zhe Li, Laurence T. Yang, Bocheng Ren, Xin Nie, Zhangyang Gao, Cheng Tan, Stan Z. Li:
MLIP: Enhancing Medical Visual Representation with Divergence Encoder and Knowledge-guided Contrastive Learning. 11704-11714
- Soumen Basu, Mayuna Gupta, Chetan Madan, Pankaj Gupta, Chetan Arora:
FocusMAE: Gallbladder Cancer Detection from Ultrasound Videos with Focused Masked Autoencoders. 11715-11725
- Xin Fan, Xiaolin Wang, Jiaxin Gao, Jia Wang, Zhongxuan Luo, Risheng Liu:
Bi-level Learning of Task-Specific Decoders for Joint Registration and One-Shot Medical Image Segmentation. 11726-11735
- Ruining Deng, Quan Liu, Can Cui, Tianyuan Yao, Jialin Yue, Juming Xiong, Lining Yu, Yifei Wu, Mengmeng Yin, Yu Wang, Shilin Zhao, Yucheng Tang, Haichun Yang, Yuankai Huo:
PrPSeg: Universal Proposition Learning for Panoramic Renal Pathology Segmentation. 11736-11746
- Xiaoyang Chen, Hao Zheng, Yuemeng Li, Yuncong Ma, Liang Ma, Hongming Li, Yong Fan:
Versatile Medical Image Segmentation Learned from Multi-Source Datasets via Model Self-Disambiguation. 11747-11756
- Oren Kraus, Kian Kenyon-Dean, Saber Saberian, Maryam Fallah, Peter McLean, Jess Leung, Vasudev Sharma, Ayla Khan, Jia Balakrishnan, Safiye Celik, Dominique Beaini, Maciej Sypetkowski, Chi Vicky Cheng, Kristen Morse, Maureen Makes, Ben Mabey, Berton Earnshaw:
Masked Autoencoders for Microscopy are Scalable Learners of Cellular Biology. 11757-11768
- Md Mostafijur Rahman, Mustafa Munir, Radu Marculescu:
EMCAD: Efficient Multi-Scale Convolutional Attention Decoding for Medical Image Segmentation. 11769-11779
- Yunkai Tang, Chengxuan Zhu, Renjie Wan, Chao Xu, Boxin Shi:
Neural Underwater Scene Representation. 11780-11789
- Mason Long Wang, Ryosuke Sawata, Samuel Clarke, Ruohan Gao, Shangzhe Wu, Jiajun Wu:
Hearing Anything Anywhere. 11790-11799
- Fan Fei, Jiajun Tang, Ping Tan, Boxin Shi:
VMINer: Versatile Multi-view Inverse Rendering with Near-and Far-field Light Sources. 11800-11809
- Heng Guo, Jieji Ren, Feishi Wang, Boxin Shi, Mingjun Ren, Yasuyuki Matsushita:
DiLiGenRT: A Photometric Stereo Dataset with Quantified Roughness and Translucency. 11810-11820
- Yufei Han, Heng Guo, Koki Fukai, Hiroaki Santo, Boxin Shi, Fumio Okura, Zhanyu Ma, Yunpeng Jia:
NeRSP: Neural 3D Reconstruction for Reflective Objects with Sparse Polarized Images. 11821-11830
- Seokjun Choi, Seungwoo Yoon, Giljoo Nam, Seungyong Lee, Seung-Hwan Baek:
Differentiable Display Photometric Stereo. 11831-11840
- Deshan Gong, Ningtao Mao, He Wang:
Bayesian Differentiable Physics for Cloth Digitalization. 11841-11851
- Fan Zhang, Shaodi You, Yu Li, Ying Fu:
Atlantis: Enabling Underwater Depth Estimation with Stable Diffusion. 11852-11861
- Mohammed Brahim, Bjoern Haefner, Zhenzhang Ye, Bastian Goldluecke, Daniel Cremers:
Sparse Views, Near Light: A Practical Paradigm for Uncalibrated Point-Light Photometric Stereo. 11862-11872
- Yuto Enyo, Ko Nishino:
Diffusion Reflectance Map: Single-Image Stochastic Inverse Rendering of Illumination and Reflectance. 11873-11883
- Nobuhiko Wakai, Satoshi Sato, Yasunori Ishii, Takayoshi Yamashita:
Deep Single Image Camera Calibration by Heatmap Regression to Recover Fisheye Images Under Manhattan World Assumption. 11884-11894
- David Stotko, Nils Wandel, Reinhard Klein:
Physics-guided Shape-from-Template: Monocular Video Perception through Neural Surrogate Models. 11895-11904
- Zongrui Li, Zhan Lu, Haojie Yan, Boxin Shi, Gang Pan, Qian Zheng, Xudong Jiang:
Spin-UP: Spin Light for Natural Light Uncalibrated Photometric Stereo. 11905-11914
- Hyomin Kim, Yucheol Jung, Seungyong Lee:
Discontinuity-preserving Normal Integration with Auxiliary Edges. 11915-11923
- Mani Ramanagopal, Sriram Narayanan, Aswin C. Sankaranarayanan, Srinivasa G. Narasimhan:
A Theory of Joint Light and Heat Transport for Lambertian Scenes. 11924-11933
- Yunshu Dai, Jianwei Fei, Fangjun Huang:
IDGuard: Robust, General, Identity-Centric POI Proactive Defense Against Face Editing Abuse. 11934-11943
- Jingwen Ye, Xinchao Wang:
Ungeneralizable Examples. 11944-11953
- Jingwen Ye, Ruonan Yu, Songhua Liu, Xinchao Wang:
Distilled Datamodel with Reverse Gradient Matching. 11954-11963



- Xuanyu Zhang, Runyi Li, Jiwen Yu, Youmin Xu, Weiqi Li, Jian Zhang:
EditGuard: Versatile Image Watermarking for Tamper Localization and Copyright Protection. 11964-11974
- Phillip Howard, Avinash Madasu, Tiep Le, Gustavo A. Lujan-Moreno, Anahita Bhiwandiwalla, Vasudev Lal:
SocialCounterfactuals: Probing and Mitigating Intersectional Social Biases in Vision-Language Models with Counterfactual Examples. 11975-11985
- Xiyuan Yang, Wenke Huang, Mang Ye:
FedAS: Bridging Inconsistency in Personalized Federated Learning. 11986-11995
- Robik Shrestha, Yang Zou, Qiuyu Chen, Zhiheng Li, Yusheng Xie, Siqi Deng:
FairRAG: Fair Human Generation via Fair Retrieval Augmentation. 11996-12005
- Hang Li, Chengzhi Shen, Philip Torr, Volker Tresp, Jindong Gu:
Self-Discovering Interpretable Diffusion Latent Directions for Responsible Text-to-Image Generation. 12006-12016
- Riddhi Chakraborty, Adrian Sletten, Michael C. Kampffmeyer:
ExMap: Leveraging Explainability Heatmaps for Unsupervised Group Robustness to Spurious Correlations. 12017-12026
- Wenqian Li, Shuran Fu, Fengrui Zhang, Yan Pang:
Data Valuation and Detections in Federated Learning. 12027-12036
- Sepehr Dehdashtian, Bashir Sadeghi, Vishnu Naresh Boddeti:
Utility-Fairness Trade-Offs and how to Find Them. 12038-12046
- Feifei Wang, Zhentao Tan, Tianyi Wei, Yue Wu, Qidong Huang:
SimAC: A Simple Anti-Customization Method for Protecting Face Privacy Against Text-to-Image Synthesis of Diffusion Models. 12047-12056
- Jun Bao, Buyu Liu, Kui Ren, Jun Yu:
GLOW: Global Layout Aware Attacks on Object Detection. 12057-12066
- Taeuk Jang, Xiaoqian Wang:
FADES: Fair Disentanglement with Sensitive Relevance. 12067-12076
- Yuhang Chen, Wenke Huang, Mang Ye:
Fair Federated Learning Under Domain Skew with Local Consistency and Domain Diversity. 12077-12086
- Youngdong Jang, Dong In Lee, MinHyuk Jang, Jong Wook Kim, Feng Yang, Sangpil Kim:
WateRF: Robust Watermarks in Radiance Fields for Protection of Copyrights. 12087-12097
- Junyuan Zhang, Shuang Zeng, Miao Zhang, Runxi Wang, Feifei Wang, Yuyin Zhou, Paul Pu Liang, Liangqiong Qu:
FLHetBench: Benchmarking Device and State Heterogeneity in Federated Learning. 12098-12108
- Jianqing Zhang, Yang Liu, Yang Hua, Jian Cao:
An Upload-Efficient Scheme for Transferring Knowledge From a Server-Side Pre-trained Generator to Clients in Heterogeneous Federated Learning. 12109-12119
- Jhon Lopez, Carlos Hinojosa, Henry Arguello, Bernard Ghanem:
Privacy-Preserving Optics for Enhancing Protection in Face De-Identification. 12120-12129
- Xiaoyang Xu, Mengda Yang, Wenzhe Yi, Ziang Li, Juan Wang, Hongxin Hu, Yong Zhuang, Yixin Liu:
A Stealthy Wrongdoer: Feature-Oriented Reconstruction Attack Against Split Learning. 12130-12139
- Fei Zhu, Zhen Cheng, Xu-Yao Zhang, Cheng-Lin Liu, Zhaoxiang Zhang:
RCL: Reliable Continual Learning for Unified Failure Detection. 12140-12150
- Hongxia Li, Wei Huang, Jingya Wang, Ye Shi:
Global and Local Prompts Cooperation via Optimal Transport for Federated Learning. 12151-12161
- Zijin Yang, Kai Zeng, Kejiang Chen, Han Fang, Weiming Zhang, Nenghai Yu:
Gaussian Shading: Provable Performance-Lossless Image Watermarking for Diffusion Models. 12162-12171
- Daniela Massiceti, Camilla Longden, Agnieszka Slowik, Samuel Wills, Martin Grayson, Cecily Morrison:
Explaining CLIP's Performance Disparities on Data from Blind/Low Vision Users. 12172-12182
- Sy-Tuyen Ho, Koh Jun Hao, Keshigeyan Chandrasegaran, Ngoc-Bao Nguyen, Ngai-Man Cheung:
Model Inversion Robustness: Can Transfer Learning Help? 12183-12193
- Gianni Franchi, Olivier Laurent, Maxence Leguérat, Andrei Bursuc, Andrea Pilzer, Angela Yao:
Make Me a BNN: A Simple Strategy for Estimating Bayesian Uncertainty from Pre-trained Models. 12194-12204
- Hui Zhang, Xingbo Dong, Yen-Lung Lai, Ying Zhou, Xiaoyan Zhang, Xingguo Lv, Zhe Jin, Xuejun Li:
Validating Privacy-Preserving Face Recognition Under a Minimum Assumption. 12205-12214
- Bin Fang, Bo Li, Shuang Wu, Shouhong Ding, Ran Yi, Lizhuang Ma:
Re-Thinking Data Availability Attacks Against Deep Neural Networks. 12215-12224
- Moreno D'Incà, Elia Peruzzo, Massimiliano Mancini, Dejia Xu, Vedit Goel, Xingqian Xu, Zhangyang Wang, Humphrey Shi, Nicu Sebe:
OpenBias: Open-Set Bias Detection in Text-to-Image Generative Models. 12225-12235
- Jinseong Park, Yujin Choi, Jaewook Lee:
In-Distribution Public Data Synthesis With Diffusion Models for Differentially Private Image Classification. 12236-12246



- Joshua C. Zhao, Ahaan Dabholkar, Atul Sharma, Saurabh Bagchi:
Leak and Learn: An Attacker's Cookbook to Train Using Leaked Data from Federated Learning. 12247-12256
- Hanwen Liu, Zhicheng Sun, Yadong Mu:
Counteracting Personalized Text-to-Image Generation with Influence Watermarks. 12257-12267
- Sungho Park, Hyeran Byun:
Fair-VPT: Fair Visual Prompt Tuning for Image Classification. 12268-12278
- Seonguk Seo, Jinkyu Kim, Geeho Kim, Bohyung Han:
Relaxed Contrastive Learning for Federated Learning. 12279-12288
- Yan Luo, Min Shi, Muhammad Osama Khan, Muhammad Muneeb Afzal, Hao Huang, Shuaihang Yuan, Yu Tian, Luo Song, Ava Kouhana, Tobias Elze, Yi Fang, Mengyu Wang:
FairCLIP: Harnessing Fairness in Vision-Language Learning. 12289-12301
- Qi Cui, Ruohan Meng, Chaohui Xu, Chip-Hong Chang:
Steganographic Passport: An Owner and User Verifiable Credential for Deep Model IP Protection Without Retraining. 12302-12311
- Q. Fan, L. Shuai:
Adaptive Hyper-graph Aggregation for Modality-Agnostic Federated Learning. 12312-12321
- Yining Wang, Junjie Sun, Chenyue Wang, Mi Zhang, Min Yang:
Navigate Beyond Shortcuts: Debiased Learning through the Lens of Neural Collapse. 12322-12331
- Jeonghoon Park, Chaeyeon Chung, Jaegul Choo:
Enhancing Intrinsic Features for Debiasing via Investigating Class-Discerning Common Attributes in Bias-Contrastive Pair. 12332-12341
- Shangqian Gao, Junyi Li, Zeyu Zhang, Yanfu Zhang, Weidong Cai, Heng Huang:
Device-Wise Federated Network Pruning. 12342-12352
- Yue Niu, Ramy E. Ali, Saurav Prakash, Salman Avestimehr:
All Rivers Run to the Sea: Private Learning with Asymmetric Flows. 12353-12362
- Xiang Li, Qianli Shen, Kenji Kawaguchi:
VA3: Virtually Assured Amplification Attack on Probabilistic Copyright Protection for Text-to-Image Generative Models. 12363-12373
- Aditya Golatkar, Alessandro Achille, Luca Zancato, Yu-Xiang Wang, Ashwin Swaminathan, Stefano Soatto:
CPR: Retrieval Augmented Generation for Copyright Protection. 12374-12384
- Geeho Kim, Jinkyu Kim, Bohyung Han:
Communication-Efficient Federated Learning with Accelerated Client Gradient. 12385-12394
- Geon Yeong Park, Chanyong Jung, Sangmin Lee, Jong Chul Ye, Sang Wan Lee:
Self-Supervised Debiasing Using Low Rank Regularization. 12395-12405
- Zhenzhong Kuang, Xiaochen Yang, Yingjie Shen, Chao Hu, Jun Yu:
Facial Identity Anonymization via Intrinsic and Extrinsic Attention Distraction. 12406-12415
- Anas Al-lahham, Muhammad Zaigham Zaheer, Nurbek Tastan, Karthik Nandakumar:
Collaborative Learning of Anomalies with Privacy (CLAP) for Unsupervised Video Anomaly Detection: A New Baseline. 12416-12425
- Yiwei Yang, Anthony Z. Liu, Robert Wolfe, Aylin Caliskan, Bill Howe:
Label-Efficient Group Robustness via Out-of-Distribution Concept Curation. 12426-12434
- Chih-Hui Ho, Kuan-Chuan Peng, Nuno Vasconcelos:
Long-Tailed Anomaly Detection with Learnable Class Names. 12435-12446
- Dingkang Yang, Kun Yang, Mingcheng Li, Shunli Wang, Shuaibing Wang, Lihua Zhang:
Robust Emotion Recognition in Context Debiasing. 12447-12457
- Mingcheng Li, Dingkang Yang, Xiao Zhao, Shuaibing Wang, Yan Wang, Kun Yang, Mingyang Sun, Dongliang Kou, Ziyun Qian, Lihua Zhang:
Correlation-Decoupled Knowledge Distillation for Multimodal Sentiment Analysis with Incomplete Modalities. 12458-12468
- Inbar Huberman-Spiegelglas, Vladimir Kulikov, Tomer Michaeli:
An Edit Friendly DDPM Noise Space: Inversion and Manipulations. 12469-12478
- Jonathan F. Carter, João Jorge, Oliver Gibson, Lionel Tarassenko:
SleepVST: Sleep Staging from Near-Infrared Video Signals using Pre-Trained Transformers. 12479-12489
- Mike Ranzinger, Greg Heinrich, Jan Kautz, Pavlo Molchanov:
AM-RADIO: Agglomerative Vision Foundation Model Reduce All Domains Into One. 12490-12500
- Jianzong Wu, Xiangtai Li, Chenyang Si, Shangchen Zhou, Jingkang Yang, Jiangning Zhang, Yining Li, Kai Chen, Yunhai Tong, Ziwei Liu, Chen Change Loy:
Towards Language-Driven Video Inpainting via Multimodal Large Language Models. 12501-12511
- Gihun Lee, Minchan Jeong, Sangmook Kim, Jaehoon Oh, Se-Young Yun:
FedSOL: Stabilized Orthogonal Learning with Proximal Restrictions in Federated Learning. 12512-12522



- Shuaibo Li, Wei Ma, Jianwei Guo, Shibiao Xu, Benchong Li, Xiaopeng Zhang:
UnionFormer: Unified-Learning Transformer with Multi-View Representation for Image Manipulation Detection and Localization. 12523-12533
- Xiang Ji, Haiyang Jiang, Yinqiang Zheng:
Motion Blur Decomposition with Cross-shutter Guidance. 12534-12543
- Yanjie Wang, Xu Zou, Luxin Yan, Sheng Zhong, Jiahuan Zhou:
SNIDA: Unlocking Few-Shot Object Detection with Non-Linear Semantic Decoupling Augmentation. 12544-12553
- Tianrun Chen, Chaotao Ding, Shangzhan Zhang, Chunyan Yu, Ying Zang, Zejian Li, Sida Peng, Lingyun Sun:
Rapid 3D Model Generation with Intuitive 3D Input. 12554-12564
- Hmrishav Bandyopadhyay, Ayan Kumar Bhunia, Pinaki Nath Chowdhury, Aneeshan Sain, Tao Xiang, Timothy M. Hospedales, Yi-Zhe Song:
SketchINR: A First Look into Sketches as Implicit Neural Representations. 12565-12574
- Jingyu Zhang, Kun Yang, Yilei Wang, Hanqi Wang, Peng Sun, Liang Song:
ERMVP: Communication-Efficient and Collaboration-Robust Multi-Vehicle Perception in Challenging Environments. 12575-12584
- Chao Zhang, Mohan Li, Ignas Budvytis, Stephan Liwicki:
DiaLoc: An Iterative Approach to Embodied Dialog Localization. 12585-12593
- Satish Kumar, Bowen Zhang, Chandrakanth Gudavalli, Connor Levenson, Lacey Hughey, Jared A. Stabach, Irene Amoke, Gordon Ojwang, Joseph Mukeka, Stephen Mwiu, Joseph Ogutu, Howard Frederick, B. S. Manjunath:
WildlifeMapper: Aerial Image Analysis for Multi-Species Detection and Identification. 12594-12604
- Muhammad Kashif Ali, Eun Woo Im, Dongjin Kim, Tae Hyun Kim:
Harnessing Meta-Learning for Improving Full-Frame Video Stabilization. 12605-12614
- Yuzheng Wang, Dingkang Yang, Zhaoyu Chen, Yang Liu, Siao Liu, Wenqiang Zhang, Lihua Zhang, Lizhe Qi:
De-Confounded Data-Free Knowledge Distillation for Handling Distribution Shifts. 12615-12625
- Hongchao Li, Jingong Chen, Aihua Zheng, Yong Wu, Yonglong Luo:
Day-Night Cross-domain Vehicle Re-identification. 12626-12635
- Mang Tik Chiu, Yuqian Zhou, Lingzhi Zhang, Zhe Lin, Connelly Barnes, Sohrab Amirghodsi, Eli Shechtman, Humphrey Shi:
Brush2Prompt: Contextual Prompt Generator for Object Inpainting. 12636-12645
- Guanqun Wang, Jiaming Liu, Chenxuan Li, Yuan Zhang, Junpeng Ma, Xinyu Wei, Kevin Zhang, Maurice Chong, Renrui Zhang, Yijiang Liu, Shanghang Zhang:
Cloud-Device Collaborative Learning for Multimodal Large Language Models. 12646-12655
- Runqi Qiao, Lan Yang, Kaiyue Pang, Honggang Zhang:
Making Visual Sense of Oracle Bones for You and Me. 12656-12665
- Zhipeng Du, Miaojing Shi, Jiankang Deng:
Boosting Object Detection with Zero-Shot Day-Night Domain Adaptation. 12666-12676
- Dongqing Wang, Tong Zhang, Alaa Abboud, Sabine Süsstrunk:
InNeRF360: Text-Guided 3D-Consistent Object Inpainting on 360° Neural Radiance Fields. 12677-12686
- Shihong Liu, Samuel Yu, Zhiqiu Lin, Deepak Pathak, Deva Ramanan:
Language Models as Black-Box Optimizers for Vision-Language Models. 12687-12697
- Zhuangzhuang Chen, Zhuonan Lai, Jie Chen, Jianqiang Li:
Mind marginal non-crack regions: Clustering-inspired representation learning for crack segmentation. 12698-12708
- Zigang Geng, Binxin Yang, Tiankai Hang, Chen Li, Shuyang Gu, Ting Zhang, Jianmin Bao, Zheng Zhang, Houqiang Li, Han Hu, Dong Chen, Baining Guo:
InstructDiffusion: A Generalist Modeling Interface for Vision Tasks. 12709-12720
- Haohan Weng, Danqing Huang, Yu Qiao, Zheng Hu, Chin-Yew Lin, Tong Zhang, C. L. Philip Chen:
Desigen: A Pipeline for Controllable Design Template Generation. 12721-12732
- Wen Yin, Jian Lou, Pan Zhou, Yulai Xie, Dan Feng, Yuhua Sun, Tailai Zhang, Lichao Sun:
Physical Backdoor: Towards Temperature-Based Backdoor Attacks in the Physical World. 12733-12743
- Su Sun, Cheng Zhao, Yuliang Guo, Ruoyu Wang, Xinyu Huang, Yingjie Victor Chen, Liu Ren:
Behind the Veil: Enhanced Indoor 3D Scene Reconstruction with Occluded Surfaces Completion. 12744-12753
- Gabriele Moreno Berton, Alex Stoken, Barbara Caputo, Carlo Masone:
EarthLoc: Astronaut Photography Localization by Indexing Earth from Space. 12754-12764
- Zeqin Yu, Jiangqun Ni, Yuzhen Lin, Haoyi Deng, Bin Li:
DiffForensics: Leveraging Diffusion Prior to Image Forgery Detection and Localization. 12765-12774
- Zhihang Dong, Xiulong Liu, Bin Chen, Pawel Polak, Peng Zhang:
MuseChat: A Conversational Music Recommendation System for Videos. 12775-12785
- Gabriele Trivigno, Carlo Masone, Barbara Caputo, Torsten Sattler:
The Unreasonable Effectiveness of Pre-Trained Features for Camera Pose Refinement. 12786-12798
- Nyeong-Ho Shin, Seon-Ho Lee, Chang-Su Kim:
Blind Image Quality Assessment Based on Geometric Order Learning. 12799-12808



- Yasiru Ranasinghe, Nithin Gopalakrishnan Nair, Wele Gedara Chaminda Bandara, Vishal M. Patel: **CrowdDiff: Multi-Hypothesis Crowd Density Estimation Using Diffusion Models.** 12809-12819
- Yichen Li, Qunwei Li, Haozhao Wang, Ruixuan Li, Wenliang Zhong, Guannan Zhang: **Towards Efficient Replay in Federated Incremental Learning.** 12820-12829
- Zhicheng Zhang, Pancheng Zhao, Eunil Park, Jufeng Yang: **MART: Masked Affective RepresenTation Learning via Masked Temporal Distribution Distillation.** 12830-12840
- Ruoqi Wang, Zhuoyang Chen, Jiayi Zhu, Qiong Luo, Feng Wang: **PolarRec: Improving Radio Interferometric Data Reconstruction Using Polar Coordinates.** 12841-12850
- Mohammed Haroon Dupty, Yanfei Dong, Sicong Leng, Guoji Fu, Yong Liang Goh, Wei Lu, Wee Sun Lee: **Constrained Layout Generation with Factor Graphs.** 12851-12860
- Feng Li, Qing Jiang, Hao Zhang, Tianhe Ren, Shilong Liu, Xueyan Zou, Huazhe Xu, Hongyang Li, Jianwei Yang, Chunyuan Li, Lei Zhang, Jianfeng Gao: **Visual in-Context Prompting.** 12861-12871
- Qiang Wang, Bingyan Liu, Yawen Li: **Traceable Federated Continual Learning.** 12872-12881
- Biqing Qi, Xinquan Chen, Junqi Gao, Dong Li, Jianxing Liu, Ligang Wu, Bowen Zhou: **Interactive Continual Learning: Fast and Slow Thinking.** 12882-12892
- Lukas Haas, Michal Skreta, Silas Alberti, Chelsea Finn: **PIGEON: Predicting Image Geolocations.** 12893-12902
- Nisarg A. Shah, Vibashan VS, Vishal M. Patel: **LQMFormer: Language-Aware Query Mask Transformer for Referring Image Segmentation.** 12903-12913
- Mu Cai, Haotian Liu, Siva Karthik Mustikovela, Gregory P. Meyer, Yuning Chai, Dennis Park, Yong Jae Lee: **ViP-LLaVA: Making Large Multimodal Models Understand Arbitrary Visual Prompts.** 12914-12923
- Ji Zhang, Shihan Wu, Lianli Gao, Heng Tao Shen, Jingkuan Song: **DePT: Decoupled Prompt Tuning.** 12924-12933
- Shangzhe Di, Weidi Xie: **Grounded Question-Answering in Long Egocentric Videos.** 12934-12943
- Qifan Yu, Juncheng Li, Longhui Wei, Liang Pang, Wentao Ye, Bosheng Qin, Siliang Tang, Qi Tian, Yueling Zhuang: **HalluciDoctor: Mitigating Hallucinatory Toxicity in Visual Instruction Data.** 12944-12953
- Jieneng Chen, Qihang Yu, Xiaohui Shen, Alan L. Yuille, Liang-Chieh Chen: **ViTamin: Designing Scalable Vision Models in the Vision-Language Era.** 12954-12966
- Ragav Sachdeva, Andrew Zisserman: **The Manga Whisperer: Automatically Generating Transcriptions for Comics.** 12967-12976
- Kanchana Ranasinghe, Satya Narayan Shukla, Omid Poursaeed, Michael S. Ryoo, Tsung-Yu Lin: **Learning to Localize Objects Improves Spatial Reasoning in Visual-LLMs.** 12977-12987
- Shubham Parashar, Zhiqiu Lin, Tian Liu, Xiangjue Dong, Yanan Li, Deva Ramanan, James Caverlee, Shu Kong: **The Neglected Tails in Vision-Language Models.** 12988-12997
- Wenxuan Wang, Tongtian Yue, Yisi Zhang, Longteng Guo, Xingjian He, Xinlong Wang, Jing Liu: **Unveiling Parts Beyond Objects: Towards Finer-Granularity Referring Expression Segmentation.** 12998-13008
- Hanoona Abdul Rasheed, Muhammad Maaz, Sahal Shaji Mullappilly, Abdelrahman M. Shaker, Salman H. Khan, Hisham Cholakkal, Rao Muhammad Anwer, Eric P. Xing, Ming-Hsuan Yang, Fahad Shahbaz Khan: **GLaMM: Pixel Grounding Large Multimodal Model.** 13009-13018
- Zeyi Sun, Ye Fang, Tong Wu, Pan Zhang, Yuhang Zang, Shu Kong, Yuanjun Xiong, Dahua Lin, Jiaqi Wang: **Alpha-CLIP: A CLIP Model Focusing on Wherever you Want.** 13019-13029
- Jiarui Xu, Xingyi Zhou, Shen Yan, Xiuye Gu, Anurag Arnab, Chen Sun, Xiaolong Wang, Cordelia Schmid: **Pixel Aligned Language Models.** 13030-13039
- Qinghao Ye, Haiyang Xu, Jiabo Ye, Ming Yan, Anwen Hu, Haowei Liu, Qi Qian, Ji Zhang, Fei Huang: **mPLUG-Owl2: Revolutionizing Multi-modal Large Language Model with Modality Collaboration.** 13040-13051
- Peng Qi, Zehong Yan, Wynne Hsu, Mong-Li Lee: **Sniffer: Multimodal Large Language Model for Explainable Out-of-Context Misinformation Detection.** 13052-13062
- Yuqi Zhang, Han Luo, Yinjie Lei: **Towards CLIP-Driven Language-Free 3D Visual Grounding via 2D-3D Relational Enhancement and Consistency.** 13063-13072
- Tongtian Yue, Jie Cheng, Longteng Guo, Xingyuan Dai, Zijia Zhao, Xingjian He, Gang Xiong, Yisheng Lv, Jing Liu: **SC-Tune: Unleashing Self-Consistent Referential Comprehension in Large Vision Language Models.** 13073-13083
- Penghao Wu, Saining Xie: **V*: Guided Visual Search as a Core Mechanism in Multimodal LLMs.** 13084-13094
- Ruozhen He, Paola Cascante-Bonilla, Ziyan Yang, Alexander C. Berg, Vicente Ordonez: **Improved Visual Grounding through Self-Consistent Explanations.** 13095-13105



- Yue Zhao, Long Zhao, Xingyi Zhou, Jialin Wu, Chun-Te Chu, Hui Miao, Florian Schroff, Hartwig Adam, Ting Liu, Boqing Gong, Philipp Krähenbühl, Liangzhe Yuan:
Distilling Vision-Language Models on Millions of Videos. 13106-13116
- Mark Hamilton, Andrew Zisserman, John R. Hershey, William T. Freeman:
Separating the "Chirp" from the "Chat": Self-supervised Visual Grounding of Sound and Language. 13117-13127
- Chang Liu, Xiangtai Li, Henghui Ding:
Referring Image Editing: Object-Level Image Editing via Referring Expressions. 13128-13138
- Liuyi Wang, Zongtao He, Ronghao Dang, Mengjiao Shen, Chengju Liu, Qijun Chen:
Vision-and-Language Navigation via Causal Learning. 13139-13150
- Fan Ma, Xiaojie Jin, Heng Wang, Yuchen Xian, Jiashi Feng, Yi Yang:
Vista-llama: Reducing Hallucination in Video Language Models via Equal Distance to Visual Tokens. 13151-13160
- Tianming Liang, Chaolei Tan, Beihao Xia, Wei-Shi Zheng, Jian-Fang Hu:
Ranking Distillation for Open-Ended Video Question Answering with Insufficient Labels. 13161-13170
- Shuyang Sun, Runjia Li, Philip Torr, Xiuye Gu, Siyang Li:
CLIP as RNN: Segment Countless Visual Concepts without Training Endeavor. 13171-13182
- Mehmet Saygin Seyfioglu, Wisdom Oluchi Ikezogwo, Fatemeh Ghezloo, Ranjay Krishna, Linda G. Shapiro:
Quilt-LLaVA: Visual Instruction Tuning by Extracting Localized Narratives from Open-Source Histopathology Videos. 13183-13192
- Yunhang Shen, Chaoyou Fu, Peixian Chen, Mengdan Zhang, Ke Li, Xing Sun, Yunsheng Wu, Shaohui Lin, Rongrong Ji:
Aligning and Prompting Everything All at Once for Universal Visual Perception. 13193-13203
- Junbin Xiao, Angela Yao, Yicong Li, Tat-Seng Chua:
Can I Trust Your Answer? Visually Grounded Video Question Answering. 13204-13214
- Yuechen Zhang, Shengju Qian, Bohao Peng, Shu Liu, Jiaya Jia:
Prompt Highlighter: Interactive Control for Multi-Modal LLMs. 13215-13224
- Geonmo Gu, Sanghyuk Chun, Wonjae Kim, Yoohoon Kang, Sangdoo Yun:
Language-only Efficient Training of Zero-shot Composed Image Retrieval. 13225-13234
- Juhong Min, Shyamal Buch, Arsha Nagrani, Minsu Cho, Cordelia Schmid:
MoRevQA: Exploring Modular Reasoning Models for Video Question Answering. 13235-13245
- Shanshan Zhong, Zhongzhan Huang, Shanghua Gao, Wushao Wen, Liang Lin, Marinka Zitnik, Pan Zhou:
Let's Think Outside the Box: Exploring Leap-of-Thought in Large Language Models with Creative Humor Generation. 13246-13257
- Zhi Gao, Yuntao Du, Xintong Zhang, Xiaojian Ma, Wenjuan Han, Song-Chun Zhu, Qing Li:
CLOVA: A Closed-Loop Visual Assistant with Tool Usage and Update. 13258-13268
- Chun Feng, Joy Hsu, Weiyu Liu, Jiajun Wu:
Naturally Supervised 3D Visual Grounding with Language-Regularized Concept Learners. 13269-13278
- Wujian Peng, Sicheng Xie, Zuyao You, Shiyi Lan, Zuxuan Wu:
Synthesize, Diagnose, and Optimize: Towards Fine-Grained Vision-Language Understanding. 13279-13288
- Difei Gao, Lei Ji, Zechen Bai, Mingyu Ouyang, Peiran Li, Dongxing Mao, Qinchen Wu, Weichen Zhang, Peiyi Wang, Xiangwu Guo, Hengxu Wang, Luowei Zhou, Mike Zheng Shou:
AssistGUI: Task-Oriented PC Graphical User Interface Automation. 13289-13298
- Bohao Li, Yuying Ge, Yixiao Ge, Guangzhi Wang, Rui Wang, Ruimao Zhang, Ying Shan:
SEED-Bench: Benchmarking Multimodal Large Language Models. 13299-13308
- Mainak Singha, Ankit Jha, Shirsha Bose, Ashwin Nair, Moloud Abdar, Biplab Banerjee:
Unknown Prompt, the only Lacuna: Unveiling CLIP's Potential for Open Domain Generalization. 13309-13319
- Tsai-Shien Chen, Aliaksandr Siarohin, Willi Menapace, Ekaterina Deyneka, Hsiang-wei Chao, Byung Eun Jeon, Yuwei Fang, Hsin-Ying Lee, Jian Ren, Ming-Hsuan Yang, Sergey Tulyakov:
Panda-70M: Captioning 70M Videos with Multiple Cross-Modality Teachers. 13320-13331
- Shuting He, Henghui Ding:
Decoupling Static and Hierarchical Motion Perception for Referring Video Segmentation. 13332-13341
- Shitian Zhao, Zhuowan Li, Yadong Lu, Alan L. Yuille, Yan Wang:
Causal-CoG: A Causal-Effect Look at Context Generation for Boosting Multi-Modal Language Models. 13342-13351
- Juil Koo, Chanho Park, Minhyuk Sung:
Posterior Distillation Sampling. 13352-13361
- Dianmo Sheng, Dongdong Chen, Zhentao Tan, Qiankun Liu, Qi Chu, Jianmin Bao, Tao Gong, Bin Liu, Shengwei Xu, Nenghai Yu:
Towards More Unified In-Context Visual Understanding. 13362-13372
- Haoquan Zhang, Ronggang Huang, Yi Xie, Huaidong Zhang:
Mask4Align: Aligned Entity Prompting with Color Masks for Multi-Entity Localization Problems. 13373-13383
- Andong Wang, Bo Wu, Sunli Chen, Zhenfang Chen, Haotian Guan, Wei-Ning Lee, Li Erran Li, Chuang Gan:
SOK-Bench: A Situated Video Reasoning Benchmark with Aligned Open-World Knowledge. 13384-13394



- Zhaohe Liao, Jiangtong Li, Li Niu, Liqing Zhang:
Align and Aggregate: Compositional Reasoning with Video Alignment and Answer Aggregation for Video Question-Answering. 13395-13404
- Xiaoke Huang, Jianfeng Wang, Yansong Tang, Zheng Zhang, Han Hu, Jiwen Lu, Lijuan Wang, Zicheng Liu:
Segment and Caption Anything. 13405-13417
- Qidong Huang, Xiaoyi Dong, Pan Zhang, Bin Wang, Conghui He, Jiaqi Wang, Dahua Lin, Weiming Zhang, Nenghai Yu:
OPERA: Alleviating Hallucination in Multi-Modal Large Language Models via Over-Trust Penalty and Retrospection-Allocation. 13418-13427
- Rongjie Li, Yu Wu, Xuming He:
Learning by Correction: Efficient Tuning Task for Zero-Shot Generative Vision-Language Reasoning. 13428-13437
- Zhihan Yu, Ruifan Li:
Revisiting Counterfactual Problems in Referring Expression Comprehension. 13438-13448
- Wei Su, Peihan Miao, Huanzhang Dou, Xi Li:
ScanFormer: Referring Expression Comprehension by Iteratively Scanning. 13449-13458
- Tsung-Han Wu, Giscard Biamby, David M. Chan, Lisa Dunlap, Ritwik Gupta, Xudong Wang, Joseph E. Gonzalez, Trevor Darrell:
See, Say, and Segment: Teaching LMMs to Overcome False Premises. 13459-13469
- Shiwei Gan, Yafeng Yin, Zhiwei Jiang, Hongkai Wen, Lei Xie, Sanglu Lu:
SignGraph: A Sign Sequence is Worth Graphs of Nodes. 13470-13479
- Yuan Gao, Kunyu Shi, Pengkai Zhu, Edouard Belval, Oren Nuriel, Srikanth Appalaraju, Shabnam Ghadar, Zhuowen Tu, Vijay Mahadevan, Stefano Soatto:
Enhancing Vision-Language Pre-Training with Rich Supervisions. 13480-13491
- Chen Wei, Chenxi Liu, Siyuan Qiao, Zhishuai Zhang, Alan L. Yuille, Jiahui Yu:
De-Diffusion Makes Text a Strong Cross-Modal Interface. 13492-13503
- Bo He, Hengduo Li, Young Kyun Jang, Menglin Jia, Xuefei Cao, Ashish Shah, Abhinav Shrivastava, Ser-Nam Lim:
MA-LMM: Memory-Augmented Large Multimodal Model for Long-Term Video Understanding. 13504-13514
- Kyle Buettner, Sina Malakouti, Xiang Lorraine Li, Adriana Kovashka:
Incorporating Geo-Diverse Knowledge into Prompting for Increased Geographical Robustness in Object Recognition. 13515-13524
- Jilan Xu, Yifei Huang, Junlin Hou, Guo Chen, Yuejie Zhang, Rui Feng, Weidi Xie:
Retrieval-Augmented Egocentric Video Captioning. 13525-13536
- Yun-Hao Cao, Kaixiang Ji, Ziyuan Huang, Chuanyang Zheng, Jiajia Liu, Jian Wang, Jingdong Chen, Ming Yang:
Towards Better Vision-Inspired Vision-Language Models. 13537-13547
- Michael Dorkenwald, Nimrod Barazani, Cees G. M. Snoek, Yuki M. Asano:
PIN: Positional Insert Unlocks Object Localisation Abilities in VLMs. 13548-13558
- Yuiga Wada, Kanta Kaneda, Daichi Saito, Komei Sugiura:
Polos: Multimodal Metric Learning from Human Feedback for Image Captioning. 13559-13568
- Chaolei Tan, Jianhuang Lai, Wei-Shi Zheng, Jian-Fang Hu:
Siamese Learning with Joint Alignment and Regression for Weakly-Supervised Video Paragraph Grounding. 13569-13580
- Reuben Tan, Ximeng Sun, Ping Hu, Jui-Hsien Wang, Hanieh Deilamsalehy, Bryan A. Plummer, Bryan Russell, Kate Saenko:
Koala: Key Frame-Conditioned Long Video-LLM. 13581-13591
- Shiyu Zhao, Long Zhao, Vijay Kumar B. G, Yumin Suh, Dimitris N. Metaxas, Manmohan Chandraker, Samuel Schulter:
Generating Enhanced Negatives for Training Language-Based Object Detectors. 13592-13602
- Kunyu Shi, Qi Dong, Luis Goncalves, Zhuowen Tu, Stefano Soatto:
Non-autoregressive Sequence-to-Sequence Vision-Language Models. 13603-13612
- Zhuowan Li, Bhavan Jasani, Peng Tang, Shabnam Ghadar:
Synthesize Step-by-Step: Tools, Templates and LLMs as Data Generators for Reasoning-Based Chart VQA. 13613-13623
- Duo Zheng, Shijia Huang, Lin Zhao, Yiwu Zhong, Liwei Wang:
Towards Learning a Generalist Model for Embodied Navigation. 13624-13634
- Aditya Kumar Singh, Dhruv Srivastava, Makarand Tapaswi:
"Previously on..." from Recaps to Story Summarization. 13635-13646
- Chaoyi Zhang, Kevin Lin, Zhengyuan Yang, Jianfeng Wang, Linjie Li, Chung-Ching Lin, Zicheng Liu, Lijuan Wang:
MM-Narrator: Narrating Long-form Videos with Multimodal In-Context Learning. 13647-13657
- Ruyang Liu, Chen Li, Yixiao Ge, Thomas H. Li, Ying Shan, Ge Li:
BT-Adapter: Video Conversation is Feasible Without Video Instruction Tuning. 13658-13667
- Xinpeng Ding, Jianhua Han, Hang Xu, Xiaodan Liang, Wei Zhang, Xiaomeng Li:
Holistic Autonomous Driving Understanding by Bird's Eye View Injected Multi-Modal Large Models. 13668-13677
- Yunze Man, Liang-Yan Gui, Yu-Xiong Wang:
Situational Awareness Matters in 3D Vision Language Reasoning. 13678-13688



- Ju-Hee Lee, Je-Won Kang:
SRTube: Video-Language Pre-Training with Action-Centric Video Tube Features and Semantic Role Labeling. 13689-13699
- Peng Jin, Ryuichi Takanobu, Wancai Zhang, Xiaochun Cao, Li Yuan:
Chat-UniVi: Unified Visual Representation Empowers Large Language Models with Image and Video Understanding. 13700-13710
- Qiyuan Dai, Sibei Yang:
Curriculum Point Prompting for Weakly-Supervised Referring Image Segmentation. 13711-13722
- Jinheng Xie, Songhe Deng, Bing Li, Haozhe Liu, Yawen Huang, Yefeng Zheng, Jürgen Schmidhuber, Bernard Ghanem, Linlin Shen, Mike Zheng Shou:
Tune-an-Ellipse: CLIP Has Potential to Find what you Want. 13723-13732
- Jiaxuan Li, Duc Minh Vo, Akihiro Sugimoto, Hideki Nakayama:
Evcap: Retrieval-Augmented Image Captioning with External Visual-Name Memory for Open-World Comprehension. 13733-13742
- Yi-Ting Hsiao, Siavash Khodadadeh, Kevin Duarte, Wei-An Lin, Hui Qu, Mingi Kwon, Ratheesh Kalarot:
Plug-and-Play Diffusion Distillation. 13743-13752
- Zihan Wang, Xiangyang Li, Jiahao Yang, Yeqi Liu, Junjie Hu, Ming Jiang, Shuqiang Jiang:
Lookahead Exploration with Neural Radiance Representation for Continuous Vision-Language Navigation. 13753-13762
- Lin Song, Yukang Chen, Shuai Yang, Xiaohan Ding, Yixiao Ge, Ying-Cong Chen, Ying Shan:
Low-Rank Approximation for Sparse Attention in Multi-Modal LLMs. 13763-13773
- Le Zhang, Rabiul Awal, Aishwarya Agrawal:
Contrasting Intra-Modal and Ranking Cross-Modal Hard Negatives to Enhance Visio-Linguistic Compositional Understanding. 13774-13784
- Chenhao Zheng, Jieyu Zhang, Aniruddha Kembhavi, Ranjay Krishna:
Iterated Learning Improves Compositionality in Large Vision-Language Models. 13785-13795
- Qiushan Guo, Shalini De Mello, Hongxu Yin, Wonmin Byeon, Ka Chun Cheung, Yizhou Yu, Ping Luo, Sifei Liu:
RegionGPT: Towards Region Understanding Vision Language Model. 13796-13806
- Tianyu Yu, Yuan Yao, Haoye Zhang, Taiwen He, Yifeng Han, Ganqu Cui, Jinyi Hu, Zhiyuan Liu, Hai-Tao Zheng, Maosong Sun:
RLHF-V: Towards Trustworthy MLLMs via Behavior Alignment from Fine-Grained Correctional Human Feedback. 13807-13816
- Junbum Cha, Wooyoung Kang, Jonghwan Mun, Byungseok Roh:
Honeybee: Locality-Enhanced Projector for Multimodal LLM. 13817-13827
- Wenjun Wu, Lingling Zhang, Jun Liu, Xi Tang, Yaxian Wang, Shaowei Wang, Qianying Wang:
E-GPS: Explainable Geometry Problem Solving via Top-Down Solver and Bottom-Up Generator. 13828-13837
- Shiyu Xuan, Qingpei Guo, Ming Yang, Shiliang Zhang:
Pink: Unveiling the Power of Referential Comprehension for Multi-modal LLMs. 13838-13848
- Zehao Xiao, Jiayi Shen, Mohammad Mahdi Derakhshani, Shengcai Liao, Cees G. M. Snoek:
Any-Shift Prompting for Generalization Over Distributions. 13849-13860
- Roy Ganz, Yair Kittenplon, Aviad Aberdam, Elad Ben-Avraham, Oren Nuriel, Shai Mazor, Ron Litman:
Question Aware Vision Transformer for Multimodal Reasoning. 13861-13871
- Sicong Leng, Hang Zhang, Guanzheng Chen, Xin Li, Shijian Lu, Chunyan Miao, Lidong Bing:
Mitigating Object Hallucinations in Large Vision-Language Models through Visual Contrastive Decoding. 13872-13882
- Neelhar Kondapaneni, Markus Marks, Manuel Knott, Rogério Guimarães, Pietro Perona:
Text-Image Alignment for Diffusion-Based Perception. 13883-13893
- Minkuk Kim, Hyeyon Bae Kim, Jinyoung Moon, Jinwoo Choi, Seong Tae Kim:
Do You Remember? Dense Video Captioning with Cross-Modal Memory Retrieval. 13894-13904
- Eric Slyman, Stefan Lee, Scott Cohen, Kushal Kafle:
FairDeDup: Detecting and Mitigating Vision-Language Fairness Disparities in Semantic Dataset Deduplication. 13905-13916
- Yuan Wang, Yali Li, Shengjin Wang:
G³-LQ: Marrying Hyperbolic Alignment with Explicit Semantic-Geometric Modeling for 3D Visual Grounding. 13917-13926
- Hritik Bansal, Yonatan Bitton, Idan Szpektor, Kai-Wei Chang, Aditya Grover:
VideoCon: Robust Video-Language Alignment via Contrast Captions. 13927-13937
- Shiyu Zhao, Samuel Schulter, Long Zhao, Zhixing Zhang, B. G. Vijay Kumar, Yumin Suh, Manmohan Chandraker, Dimitris N. Metaxas:
Taming Self-Training for Open-Vocabulary Object Detection. 13938-13947
- Chull Hwan Song, Taebaek Hwang, Jooyoung Yoon, Shunghyun Choi, Yeong Hyeon Gu:
SyncMask: Synchronized Attentional Masking for Fashion-centric Vision-Language Pretraining. 13948-13957



- Chuang Lin, Yi Jiang, Lizhen Qu, Zehuan Yuan, Jianfei Cai:
Generative Region-Language Pretraining for Open-Ended Object Detection. 13958-13968
- Shaowei Wang, Lingling Zhang, Longji Zhu, Tao Qin, Kim-Hui Yap, Xinyu Zhang, Jun Liu:
CoG-DQA: Chain-of-Guiding Learning with Large Language Models for Diagram Question Answering. 13969-13979
- Junwen He, Yifan Wang, Lijun Wang, Huchuan Lu, Jun-Yan He, Jin-Peng Lan, Bin Luo, Xuansong Xie:
Multi-Modal Instruction Tuned LLMs with Fine-Grained Visual Perception. 13980-13990
- Fei Ni, Jianye Hao, Shiguang Wu, Longxin Kou, Jiajun Liu, Yan Zheng, Bin Wang, Yuzheng Zhuang:
Generate Subgoal Images Before Act: Unlocking the Chain-of-Thought Reasoning in Diffusion Model for Robot Manipulation with Multimodal Prompts. 13991-14000
- Linfeng Yuan, Miao Jing Shi, Zijie Yue, Qijun Chen:
LoSh: Long-Short Text Joint Prediction Network for Referring Video Object Segmentation. 14001-14010
- Haran Raajesh, Naveen Reddy Desanur, Zeeshan Khan, Makarand Tapaswi:
MICap: A Unified Model for Identity-Aware Movie Descriptions. 14011-14021
- Qiying Yu, Quan Sun, Xiaosong Zhang, Yufeng Cui, Fan Zhang, Yue Cao, Xinlong Wang, Jingjing Liu:
CapsFusion: Rethinking Image-Text Data at Scale. 14022-14032
- Yunhao Ge, Xiaohui Zeng, Jacob Samuel Huffman, Tsung-Yi Lin, Ming-Yu Liu, Yin Cui:
Visual Fact Checker: Enabling High-Fidelity Detailed Caption Generation. 14033-14042
- Mamshad Nayeem Rizve, Fan Fei, Jayakrishnan Unnikrishnan, Son Tran, Benjamin Z. Yao, Belinda Zeng, Mubarak Shah, Trishul Chilimbi:
VidLA: Video-Language Alignment at Scale. 14043-14055
- Xiangxi Shi, Zhonghua Wu, Stefan Lee:
Viewpoint-Aware Visual Grounding in 3D Scenes. 14056-14065
- Jiawei Yao, Qi Qian, Juhua Hu:
Multi-Modal Proxy Learning Towards Personalized Visual Multiple Clustering. 14066-14075
- Shraman Pramanick, Guangxing Han, Rui Hou, Sayan Nag, Ser-Nam Lim, Nicolas Ballas, Qifan Wang, Rama Chellappa, Amjad Almahairi:
Jack of All Tasks, Master of Many: Designing General-purpose Coarse-to-Fine Vision-Language Model. 14076-14088
- Bo Zou, Chao Yang, Yu Qiao, Chengbin Quan, Youjian Zhao:
LLaMA-Excitor: General Instruction Tuning via Indirect Feature Interaction. 14089-14099
- Zequn Zeng, Yan Xie, Hao Zhang, Chiyu Chen, Bo Chen, Zhengjue Wang:
MeaCap: Memory-Augmented Zero-shot Image Captioning. 14100-14110
- Yanjun Sun, Yue Qiu, Mariia Khan, Fumiya Matsuzawa, Kenji Iwata:
The STV chrono Dataset: Towards Continuous Change Recognition in Time. 14111-14120
- Chengjian Feng, Yujie Zhong, Zequn Jie, Weidi Xie, Lin Ma:
InstaGen: Enhancing Object Detection by Training on Synthetic Dataset. 14121-14130
- Chun-Peng Chang, Shaoxiang Wang, Alain Pagani, Didier Stricker:
MiKASA: Multi-Key-Anchor & Scene-Aware Transformer for 3D Visual Grounding. 14131-14140
- Yunan Zeng, Yan Huang, Jinjin Zhang, Zequn Jie, Zhenhua Chai, Liang Wang:
Investigating Compositional Challenges in Vision-Language Models for Visual Grounding. 14141-14151
- Han Qiu, Jiaxing Huang, Peng Gao, Lewei Lu, Xiaoqin Zhang, Shijian Lu:
Masked AutoDecoder is Effective Multi-Task Vision Generalist. 14152-14161
- Adilbek Karmanov, Dayan Guan, Shijian Lu, Abdulmotaleb El-Saddik, Eric P. Xing:
Efficient Test-Time Adaptation of Vision-Language Models. 14162-14171
- Adrian Bulat, Yassine Ouali, Georgios Tzimiropoulos:
FFF: Fixing Flawed Foundations in contrastive pre-training results in very strong Vision-Language models. 14172-14182
- Sebastian Koch, Narunas Vaskevicius, Mirco Colosi, Pedro Hermosilla, Timo Ropinski:
Open3DSG: Open-Vocabulary 3D Scene Graphs from Point Clouds with Queryable Objects and Open-Set Relationships. 14183-14193
- Shenshen Bu, Taiji Li, Yuedong Yang, Zhiming Dai:
Instance-level Expert Knowledge and Aggregate Discriminative Attention for Radiology Report Generation. 14194-14204
- Jialin Wu, Xia Hu, Yaqing Wang, Bo Pang, Radu Soricut:
Omni-SMoLA: Boosting Generalist Multimodal Models with Soft Mixture of Low-Rank Experts. 14205-14215
- Sepehr Sameni, Kushal Kafle, Hao Tan, Simon Jenni:
Building Vision-Language Models on Solid Foundations with Masked Distillation. 14216-14226
- Yichi Zhang, Ziqiao Ma, Xiaofeng Gao, Suhaila Shakiah, Qiaozi Gao, Joyce Chai:
Groundhog Grounding Large Language Models to Holistic Segmentation. 14227-14238
- Yangyi Chen, Karan Sikka, Michael Cogswell, Heng Ji, Ajay Divakaran:
DRESS : Instructing Large Vision-Language Models to Align and Interact with Humans via Natural Language Feedback. 14239-14250



- Yicong Li, Na Zhao, Junbin Xiao, Chun Feng, Xiang Wang, Tat-Seng Chua:
LASO: Language-Guided Affordance Segmentation on 3D Object. 14251-14260
- Sai Wang, Yutian Lin, Yu Wu:
Omni-Q: Omni-Directional Scene Understanding for Unsupervised Visual Grounding. 14261-14270
- Bin Huang, Xin Wang, Hong Chen, Zihan Song, Wenwu Zhu:
VTimeLLM: Empower LLM to Grasp Video Moments. 14271-14280
- Wenyi Hong, Weihan Wang, Qingsong Lv, Jiazheng Xu, Wenmeng Yu, Junhui Ji, Yan Wang, Zihan Wang, Yuxiao Dong, Ming Ding, Jie Tang:
CogAgent: A Visual Language Model for GUI Agents. 14281-14290
- Sijie Cheng, Zhicheng Guo, Jingwen Wu, Kechen Fang, Peng Li, Huaping Liu, Yang Liu:
EgoThink: Evaluating First-Person Perspective Thinking Capability of Vision-Language Models. 14291-14302
- Alessandro Favero, Luca Zancato, Matthew Trager, Siddharth Choudhary, Pramuditha Perera, Alessandro Achille, Ashwin Swaminathan, Stefano Soatto:
Multi-Modal Hallucination Control by Visual Information Grounding. 14303-14312
- Shuhuai Ren, Linli Yao, Shicheng Li, Xu Sun, Lu Hou:
TimeChat: A Time-sensitive Multimodal Large Language Model for Long Video Understanding. 14313-14323
- Sixing Yan, William K. Cheung, Ivor W. Tsang, Keith Chin, Terence M. Tong, Ka Chun Cheung, Simon See:
AHIVE: Anatomy-Aware Hierarchical Vision Encoding for Interactive Radiology Report Retrieval. 14324-14333
- Mayug Maniparambil, Raiymbek Akshulakov, Yasser Abdelaziz Dahou Djilali, Mohamed El Amine Seddik, Sanath Narayan, Karttikeya Mangalam, Noel E. O'Connor:
Do Vision and Language Encoders Represent the World Similarly? 14334-14343
- Zaid Khan, Vijay Kumar B. G, Samuel Schulter, Yun Fu, Manmohan Chandraker:
Self-Training Large Language Models for Improved Visual Program Synthesis With Visual Reinforcement. 14344-14353
- Khoi Pham, Chuong Huynh, Ser-Nam Lim, Abhinav Shrivastava:
Composing Object Relations and Attributes for Image-Text Matching. 14354-14363
- Zeyu Han, Fangrui Zhu, Qianru Lao, Huaizu Jiang:
Zero-Shot Referring Expression Comprehension via Structural Similarity Between Images and Captions. 14364-14375
- Tianrui Guan, Fuxiao Liu, Xiyang Wu, Ruiqi Xian, Zongxia Li, Xiaoyu Liu, Xijun Wang, Lichang Chen, Furong Huang, Yaser Yacoob, Dinesh Manocha, Tianyi Zhou:
Hallusionbench: An Advanced Diagnostic Suite for Entangled Language Hallucination and Visual Illusion in Large Vision-Language Models. 14375-14385
- Pinelopi Papalampidi, Skanda Koppula, Shreya Pathak, Justin Chiu, Joe Heyward, Viorica Patraucean, Jiajun Shen, Antoine Miech, Andrew Zisserman, Aida Nematzadeh:
A Simple Recipe for Contrastively Pre-Training Video-First Encoders Beyond 16 Frames. 14386-14397
- Quan Sun, Yufeng Cui, Xiaosong Zhang, Fan Zhang, Qiying Yu, Yueze Wang, Yongming Rao, Jingjing Liu, Tiejun Huang, Xinlong Wang:
Generative Multimodal Models are In-Context Learners. 14398-14409
- Pratyusha Sharma, Tamar Rott Shaham, Manel Baradad, Adrián Rodríguez-Muñoz, Shivam Duggal, Phillip Isola, Antonio Torralba, Stephanie Fu:
A Vision Check-up for Language Models. 14410-14419
- Chancharik Mitra, Brandon Huang, Trevor Darrell, Roei Herzig:
Compositional Chain-of-Thought Prompting for Large Multimodal Models. 14420-14431
- Xi Chen, Josip Djolonga, Piotr Padlewski, Basil Mustafa, Soravit Changpinyo, Jialin Wu, Carlos Riquelme Ruiz, Sebastian Goodman, Xiao Wang, Yi Tay, Siamak Shakeri, Mostafa Dehghani, Daniel Salz, Mario Lucic, Michael Tschannen, Arsha Nagrani, Hexiang Hu, Mandar Joshi, Bo Pang, Ceslee Montgomery, Paulina Pietrzyk, Marvin Ritter, A. J. Piergiovanni, Matthias Minderer, Filip Pavetic, Austin Waters, Gang Li, Ibrahim Alabdulmohsin, Lucas Beyer, Julien Amelot, Kenton Lee, Andreas Peter Steiner, Yang Li, Daniel Keysers, Anurag Arnab, Yuanzhong Xu, Keran Rong, Alexander Kolesnikov, Mojtaba Seyedhosseini, Anelia Angelova, Xiaohua Zhai, Neil Houlsby, Radu Soricut:
On Scaling Up a Multilingual Vision and Language Model. 14432-14444
- Jihyung Kil, Chan Hee Song, Boyuan Zheng, Xiang Deng, Yu Su, Wei-Lun Chao:
Dual-View Visual Contextualization for Web Navigation. 14445-14454
- Boyuan Chen, Zhuo Xu, Sean Kirmani, Brian Ichter, Dorsa Sadigh, Leonidas J. Guibas, Fei Xia:
SpatialVLM: Endowing Vision-Language Models with Spatial Reasoning Capabilities. 14455-14465
- Nirat Saini, Khoi Pham, Abhinav Shrivastava:
Beyond Seen Primitive Concepts and Attribute-Object Compositional Learning. 14466-14476
- Gang Zhang, Junnan Chen, Guohuan Gao, Jianmin Li, Si Liu, Xiaolin Hu:
SAFDNet: A Simple and Effective Network for Fully Sparse 3D Object Detection. 14477-14486
- Ben Agro, Quinlan Sykora, Sergio Casas, Thomas Gilles, Raquel Urtasun:
UnO: Unsupervised Occupancy Fields for Perception and Forecasting. 14487-14496
- Gen Li, Kaifeng Zhao, Siwei Zhang, Xiaozhong Lyu, Mihai Dusmanu, Yan Zhang, Marc Pollefeys, Siyu Tang:
EgoGen: An Egocentric Synthetic Data Generator. 14497-14509



- Yuhang Shen, Huiyu Wang, Xitong Yang, Matt Feiszli, Ehsan Elhamifar, Lorenzo Torresani, Effrosyni Mavroudi: **Learning to Segment Referred Objects from Narrated Egocentric Videos.** 14510-14520
- Xunjiang Gu, Guanyu Song, Igor Gilitschenski, Marco Pavone, Boris Ivanovic: **Producing and Leveraging Online Map Uncertainty in Trajectory Prediction.** 14521-14530
- Alexandros Delitzas, Ayça Takmaz, Federico Tombari, Robert W. Sumner, Marc Pollefeys, Francis Engelmann: **SceneFun3D: Fine-Grained Functionality and Affordance Understanding in 3D Scenes.** 14531-14542
- Paul Roetzer, Florian Bernard: **SpiderMatch: 3D Shape Matching with Global Optimality and Geometric Consistency.** 14543-14553
- Anh-Quan Cao, Angela Dai, Raoul de Charette: **PaSCo: Urban 3D Panoptic Scene Completion with Uncertainty Awareness.** 14554-14564
- Tzofi Klinghoffer, Xiaoyu Xiang, Siddharth Somasundaram, Yuchen Fan, Christian Richardt, Ramesh Raskar, Rakesh Ranjan: **PlatoNeRF: 3D Reconstruction in Plato's Cave via Single-View Two-Bounce Lidar.** 14565-14574
- Feng Yu, Teng Zhang, Gilad Lerman: **A Subspace-Constrained Tyler's Estimator and its Applications to Structure from Motion.** 14575-14584
- Sangmin Lee, Bolin Lai, Fiona Ryan, Bikram Boote, James M. Rehg: **Modeling Multimodal Social Interactions: New Challenges and Baselines with Densely Aligned Representations.** 14585-14595
- Ling Gao , Daniel Gehrig, Hang Su, Davide Scaramuzza, Laurent Kneip: **An N-Point Linear Solver for Line and Motion Estimation with Event Cameras.** 14596-14605
- Siwei Zhang, Bharat Lal Bhatnagar, Yuanlu Xu, Alexander Winkler, Petr Kadlec, Siyu Tang, Federica Bogo: **RoHM: Robust Human Motion Reconstruction via Diffusion.** 14606-14617
- Ming Xu, Stephen Gould: **Temporally Consistent Unbalanced Optimal Transport for Unsupervised Action Segmentation.** 14618-14627
- Jinglin Xu, Sibo Yin, Guohao Zhao, Zishuo Wang, Yuxin Peng: **FineParser: A Fine-Grained Spatio-Temporal Action Parser for Human-Centric Action Quality Assessment.** 14628-14637
- Kewei Wang, Yizheng Wu, Jun Cen, Zhiyu Pan, Xingyi Li, Zhe Wang, Zhiguo Cao, Guosheng Lin: **Self-Supervised Class-Agnostic Motion Prediction with Spatial and Temporal Consistency Regularizations.** 14638-14647
- Youquan Liu, Lingdong Kong, Xiaoyang Wu, Runnan Chen, Xin Li, Liang Pan, Ziwei Liu, Yuexin Ma: **Multi-Space Alignments Towards Universal LiDAR Segmentation.** 14648-14661
- Jiazhi Yang, Shenyuan Gao, Yihang Qiu, Li Chen, Tianyu Li, Bo Dai, Kashyap Chitta, Penghao Wu, Jia Zeng, Ping Luo, Jun Zhang, Andreas Geiger, Yu Qiao, Hongyang Li: **Generalized Predictive Model for Autonomous Driving.** 14662-14672
- Zetong Yang, Li Chen, Yanan Sun, Hongyang Li: **Visual Point Cloud Forecasting Enables Scalable Autonomous Driving.** 14673-14684
- Jenny Seidenschwarz, Aljosa Osep, Francesco Ferroni, Simon Lucey, Laura Leal-Taixé: **SeMoLi: What Moves Together Belongs Together.** 14685-14694
- Mingfu Liang, Jong-Chyi Su, Samuel Schulter, Sparsh Garg, Shiyu Zhao, Ying Wu, Manmohan Chandraker: **AIDE: An Automatic Data Engine for Object Detection in Autonomous Driving.** 14695-14706
- Xin Zhou, Dingkang Liang, Wei Xu, Xingkui Zhu, Yihan Xu, Zhikang Zou, Xiang Bai: **Dynamic Adapter Meets Prompt Tuning: Parameter-Efficient Transfer Learning for Point Cloud Analysis.** 14707-14717
- Wenjie Wang, Yehao Lu, Guangcong Zheng, Shuigen Zhan, Xiaoqing Ye, Zichang Tan, Jingdong Wang, Gaoang Wang, Xi Li: **BEVSpread: Spread Voxel Pooling for Bird's-Eye-View Representation in Vision-Based Roadside 3D Object Detection.** 14718-14727
- Simon Doll, Niklas Hanselmann, Lukas Schneider, Richard Schulz, Marius Cordts, Markus Enzweiler, Hendrik P. A. Lensch: **Dualad: Disentangling the Dynamic and Static World for End-to-End Driving.** 14728-14737
- Haoxi Ran, Vitor Guizilini, Yue Wang: **Towards Realistic Scene Generation with LiDAR Diffusion Models.** 14738-14748
- Yuqi Wang, Jiawei He, Lue Fan, Hongxin Li, Yuntao Chen, Zhaoxiang Zhang: **Driving Into the Future: Multiview Visual Forecasting and Planning with World Model for Autonomous Driving.** 14749-14759
- Chenbin Pan, Burhaneddin Yaman, Tommaso Nesti, Abhirup Mallik, Alessandro Gabriele Allievi, Senem Velipasalar, Liu Ren: **VLP: Vision Language Planning for Autonomous Driving.** 14760-14769
- Lucas Nunes, Rodrigo Marcuzzi, Benedikt Mersch, Jens Behley, Cyrill Stachniss: **Scaling Diffusion Models to Real-World 3D LiDAR Scene Completion.** 14770-14780
- Haimei Zhao, Jing Zhang, Zhuo Chen, Shanshan Zhao, Dacheng Tao: **UniMix: Towards Domain Adaptive and Generalizable LiDAR Semantic Segmentation in Adverse Weather.** 14781-14791



- Song Wang, Jiawei Yu, Wentong Li, Wenyu Liu, Xiaolu Liu, Junbo Chen, Jianke Zhu:
Not All Voxels are Equal: Hardness-Aware Semantic Scene Completion with Self-Distillation. 14792-14801
- Haichao Zhang, Yi Xu, Hongsheng Lu, Takayuki Shimizu, Yun Fu:
OOSTraj: Out-of-Sight Trajectory Prediction With Vision-Positioning Denoising. 14802-14811
- Xiaolu Liu, Song Wang, Wentong Li, Ruizi Yang, Junbo Chen, Jianke Zhu:
MGMap: Mask-Guided Learning for Online Vectorized HD Map Construction. 14812-14821
- Di Wen, Haoran Xu, Zhaocheng He, Zhe Wu, Guang Tan, Peixi Peng:
Density-Adaptive Model Based on Motif Matrix for Multi-Agent Trajectory Prediction. 14822-14832
- Yining Shi, Kun Jiang, Ke Wang, Jiusi Li, Yunlong Wang, Mengmeng Yang, Diange Yang:
StreamingFlow: Streaming Occupancy Forecasting with Asynchronous Multi-modal Data Streams via Neural Ordinary Differential Equation. 14833-14842
- Shan Wang, Chuong Nguyen, Jiawei Liu, Yanhao Zhang, Sundaram Muthu, Fahira Afzal Maken, Kaihao Zhang, Hongdong Li:
View from Above: Orthogonal-View Aware Cross-View Localization. 14843-14852
- Zetong Yang, Zhiding Yu, Christopher B. Choy, Renhao Wang, Anima Anandkumar, José M. Álvarez:
Improving Distant 3D Object Detection Using 2D Box Supervision. 14853-14863
- Zhiqi Li, Zhiding Yu, Shiyi Lan, Jiahua Li, Jan Kautz, Tong Lu, José M. Álvarez:
Is Ego Status All You Need for Open-Loop End-to-End Autonomous Driving? 14864-14873
- Mozhgan Pourkeshavarz, Junrui Zhang, Amir Rasouli:
CaDeT: A Causal Disentanglement Approach for Robust Trajectory Prediction in Autonomous Driving. 14874-14884
- Mozhgan Pourkeshavarz, Mohammad Sabokrou, Amir Rasouli:
Adversarial Backdoor Attack by Naturalistic Data Poisoning on Trajectory Prediction in Autonomous Driving. 14885-14894
- Adam Tonderski, Carl Lindström, Georg Hess, William Ljungbergh, Lennart Svensson, Christoffer Petersson:
NeuRAD: Neural Rendering for Autonomous Driving. 14895-14904
- Junbo Yin, Jianbing Shen, Runnan Chen, Wei Li, Ruigang Yang, Pascal Frossard, Wenguan Wang:
IS-Fusion: Instance-Scene Collaborative Fusion for Multimodal 3D Object Detection. 14905-14915
- Tuo Feng, Wenguan Wang, Fan Ma, Yi Yang:
LSK3DNet: Towards Effective and Efficient 3D Perception with Large Sparse Kernels. 14916-14927
- Zhiwei Lin, Zhe Liu, Zhongyu Xia, Xinhao Wang, Yongtao Wang, Shengxiang Qi, Yang Dong, Nan Dong, Le Zhang, Ce Zhu:
RCBEVDet: Radar-Camera Fusion in Bird's Eye View for 3D Object Detection. 14928-14937
- Kuan-Chih Huang, Weijie Lyu, Ming-Hsuan Yang, Yi-Hsuan Tsai:
PTT: Point-Trajectory Transformer for Efficient Temporal 3D Object Detection. 14938-14947
- Boyi Li, Yue Wang, Jiageng Mao, Boris Ivanovic, Sushant Veer, Karen Leung, Marco Pavone:
Driving Everywhere with Large Language Model Policy Adaptation. 14948-14957
- Yan Xia, Letian Shi, Zifeng Ding, João F. Henriques, Daniel Cremers:
Text2Loc: 3D Point Cloud Localization from Natural Language. 14958-14967
- Hai Wu, Shijia Zhao, Xun Huang, Chenglu Wen, Xin Li, Cheng Wang:
Commonsense Prototype for Outdoor Unsupervised 3D Object Detection. 14968-14977
- Hanshi Wang, Zhipeng Zhang, Jin Gao, Weiming Hu:
A-Teacher: Asymmetric Network for 3D Semi-Supervised Object Detection. 14978-14987
- Norman Mu, Jingwei Ji, Zhenpei Yang, Nate Harada, Haotian Tang, Kan Chen, Charles R. Qi, Runzhou Ge, Kratarth Goel, Zoey Yang, Scott Ettinger, Rami Al-Rfou, Dragomir Anguelov, Yin Zhou:
MoST: Multi-modality Scene Tokenization for Motion Prediction. 14988-14999
- Jimuyang Zhang, Zanming Huang, Arijit Ray, Eshed Ohn-Bar:
Feedback-Guided Autonomous Driving. 15000-15011
- Yiduo Hao, Sohrab Madani, Junfeng Guan, Mohammed Alloulah, Saurabh Gupta, Haitham Hassanieh:
Bootstrapping Autonomous Driving Radars with Self-Supervised Learning. 15012-15023
- Ryoma Yataka, Pu Wang, Petros Boufounos, Ryuhei Takahashi:
SIRA: Scalable Inter-Frame Relation and Association for Radar Perception. 15024-15034
- Pin Tang, Zhongdao Wang, Guoqing Wang, Jilai Zheng, Xiangxuan Ren, Bailan Feng, Chao Ma:
SparseOcc: Rethinking Sparse Latent Representation for Vision-Based Semantic Occupancy Prediction. 15035-15044
- Wen Li, Yuyang Yang, Shangshu Yu, Guosheng Hu, Chenglu Wen, Ming Cheng, Cheng Wang:
DiffLoc: Diffusion Model for Outdoor LiDAR Localization. 15045-15054
- Alexander Gambashidze, Aleksandr Dadukin, Maksim Golyadkin, Maria Razzhivina, Ilya Makarov:
Weak-to-Strong 3D Object Detection with X-Ray Distillation. 15055-15064
- Daehee Park, Jaeseok Jeong, Sung-Hoon Yoon, Jaewoo Jeong, Kuk-Jin Yoon:
T4P: Test-Time Training of Trajectory Prediction via Masked Autoencoder and Actor-Specific Token Memory. 15065-15076



- Yuxi Wei, Zi Wang, Yifan Lu, Chenxin Xu, Changxing Liu, Hao Zhao, Siheng Chen, Yanfeng Wang: **Editable Scene Simulation for Autonomous Driving via Collaborative LLM-Agents.** 15077-15087
- Lei Lai, Eshed Ohn-Bar, Sanjay Arora, John Seon Keun Yi: **Uncertainty-Guided Never-Ending Learning to Drive.** 15088-15098
- Kaituo Feng, Changsheng Li, Dongchun Ren, Ye Yuan, Guoren Wang: **On the Road to Portability: Compressing End-to-End Motion Planner for Autonomous Driving.** 15099-15108
- Jiuming Liu, Guangming Wang, Weicai Ye, Chaokang Jiang, Jinru Han, Zhe Liu, Guofeng Zhang, Dalong Du, Hesheng Wang: **DifFlow3D: Toward Robust Uncertainty-Aware Scene Flow Estimation with Iterative Diffusion-Based Refinement.** 15109-15119
- Hao Shao, Yuxuan Hu, Letian Wang, Guanglu Song, Steven L. Waslander, Yu Liu, Hongsheng Li: **LMDrive: Closed-Loop End-to-End Driving with Large Language Models.** 15120-15130
- Quentin Herau, Nathan Piasco, Moussâb Bennehar, Luis Roldao, Dzmitry Tsishkou, Cyrille Migniot, Pascal Vasseur, Cédric Demonceaux: **SOAC: Spatio-Temporal Overlap-Aware Multi-Sensor Calibration using Neural Radiance Fields.** 15131-15140
- Yunsheng Ma, Can Cui, Xu Cao, Wenqian Ye, Peiran Liu, Juanwu Lu, Amr Abdelraouf, Rohit Gupta, Kyungtae Han, Aniket Bera, James M. Rehg, Ziran Wang: **LaMPilot: An Open Benchmark Dataset for Autonomous Driving with Language Model Programs.** 15141-15151
- Prashant Kumar, Kshitij Madhav Bhat, Vedang Bhupesh Shenvi Nadkarni, Prem Kalra: **GLiDR: Topologically Regularized Graph Generative Network for Sparse LiDAR Point Clouds.** 15152-15161
- Yujeong Chae, Hyeonseong Kim, Kuk-Jin Yoon: **Towards Robust 3D Object Detection with LiDAR and 4D Radar Fusion in Various Weather Conditions.** 15162-15172
- Chaokang Jiang, Guangming Wang, Jiuming Liu, Hesheng Wang, Zhuang Ma, Zhenqiang Liu, Zhujin Liang, Yi Shan, Dalong Du: **3DSFLabelling: Boosting 3D Scene Flow Estimation by Pseudo Auto-Labelling.** 15173-15183
- Shuxiao Ding, Lukas Schneider, Marius Cordts, Juergen Gall: **ADA-Track: End-to-End Multi-Camera 3D Multi-Object Tracking with Alternating Detection and Association.** 15184-15194
- Loïck Chambon, Éloi Zablocki, Mickaël Chen, Florent Bartoccioni, Patrick Pérez, Matthieu Cord: **PointBeV: A Sparse Approach to BeV Predictions.** 15195-15204
- Jinlong Li, Baolu Li, Zhengzhong Tu, Xinyu Liu, Qing Guo, Felix Juefei-Xu, Runsheng Xu, Hongkai Yu: **Light the Night: A Multi-Condition Diffusion Framework for Unpaired Low-Light Enhancement in Autonomous Driving.** 15205-15215
- Chenbin Pan, Burhaneddin Yaman, Senem Velipasalar, Liu Ren: **CLIP-BEVFormer: Enhancing Multi-View Image-Based BEV Detector with Ground Truth Flow.** 15216-15225
- Yi Xu, Yun Fu: **Adapting to Length Shift: FlexiLength Network for Trajectory Prediction.** 15226-15237
- Honghui Yang, Sha Zhang, Di Huang, Xiaoyang Wu, Haoyi Zhu, Tong He, Shixiang Tang, Hengshuang Zhao, Qibo Qiu, Binbin Lin, Xiaofei He, Wanli Ouyang: **UniPAD: A Universal Pre-Training Paradigm for Autonomous Driving.** 15238-15250
- Sungjune Kim, Hyung-Gun Chi, Hyerin Lim, Karthik Ramani, Jinkyu Kim, Sangpil Kim: **Higher-order Relational Reasoning for Pedestrian Trajectory Prediction.** 15251-15260
- Xiaolong Tang, Meina Kan, Shiguang Shan, Zhilong Ji, Jinfeng Bai, Xilin Chen: **HPNet: Dynamic Trajectory Forecasting with Historical Prediction Attention.** 15261-15270
- Bochun Yang, Zijun Li, Wen Li, Zhipeng Cai, Chenglu Wen, Yu Zang, Matthias Müller, Cheng Wang: **LiSA: LiDAR Localization with Semantic Awareness.** 15271-15280
- Yang Zhou, Hao Shao, Letian Wang, Steven L. Waslander, Hongsheng Li, Yu Liu: **SmartRefine: A Scenario-Adaptive Refinement Framework for Efficient Motion Prediction.** 15281-15290
- Zhanwei Zhang, Minghao Chen, Shuai Xiao, Liang Peng, Hengjia Li, Binbin Lin, Ping Li, Wenxiao Wang, Boxi Wu, Deng Cai: **Pseudo Label Refinery for Unsupervised Domain Adaptation on Cross-Dataset 3D Object Detection.** 15291-15300
- Shixin Hong, Yu Liu, Zhi Li, Shaohui Li, You He: **Multi-Agent Collaborative Perception via Motion-Aware Robust Communication Network.** 15301-15310
- Xiaopei Wu, Yuenan Hou, Xiaoshui Huang, Binbin Lin, Tong He, Xinge Zhu, Yuexin Ma, Boxi Wu, Haifeng Liu, Deng Cai, Wanli Ouyang: **TASeg: Temporal Aggregation Network for LiDAR Semantic Segmentation.** 15311-15320
- Qiming Xia, Wei Ye, Hai Wu, Shijia Zhao, Leyuan Xing, Xun Huang, Jinhao Deng, Xin Li, Chenglu Wen, Cheng Wang: **HINTED: Hard Instance Enhanced Detector with Mixed-Density Feature Fusion for Sparsely-Supervised 3D Object Detection.** 15321-15330
- Haonan Zhang, Longjun Liu, Yuqi Huang, Zhao Yang, Xinyu Lei, Bihan Wen: **CaKDP: Category-Aware Knowledge Distillation and Pruning Framework for Lightweight 3D Object Detection.** 15331-15341
- Brian Yang, Huangyuan Su, Nikolaos Gkanatsios, Tsung-Wei Ke, Ayush Jain, Jeff G. Schneider, Katerina Fragkiadaki: **Diffusion-ES: Gradient-Free Planning with Diffusion for Autonomous and Instruction-Guided Driving.** 15342-15353



- Bin Yang, Patrick Pfreundschuh, Roland Siegwart, Marco Hutter, Peyman Moghadam, Vaishakh Patil: **TULIP: Transformer for Upsampling of LiDAR Point Clouds.** 15354-15364
- Hugh Blayney, Hanlin Tian, Hamish Scott, Nils Goldbeck, Chess Stetson, Panagiotis Angeloudis: **Bézier Everywhere All at Once: Learning Drivable Lanes as Bézier Graphs.** 15365-15374
- Anush Kumar, Fahim Mannan, Omid Hosseini Jafari, Shile Li, Felix Heide: **Flow-Guided Online Stereo Rectification for Wide Baseline Stereo.** 15375-15385
- Ke Guo, Zhenwei Miao, Wei Jing, Weiwei Liu, Weizi Li, Dayang Hao, Jia Pan: **LASIL: Learner-Aware Supervised Imitation Learning For Long-Term Microscopic Traffic Simulation.** 15386-15395
- Yi Zhou, Hui Zhang, Jiaqian Yu, Yifan Yang, Sangil Jung, Seung-In Park, ByungIn Yoo: **HIMap: Hybrid Representation Learning for End-to-end Vectorized HD Map Construction.** 15396-15406
- Oded Bialer, Yuval Haitman: **RadSimReal: Bridging the Gap Between Synthetic and Real Data in Radar Object Detection With Simulation.** 15407-15416
- Xingguang Zhong, Yue Pan, Cyrill Stachniss, Jens Behley: **3D LiDAR Mapping in Dynamic Environments Using a 4D Implicit Neural Representation.** 15417-15427
- Juanwu Lu, Can Cui, Yunsheng Ma, Aniket Bera, Ziran Wang: **Quantifying Uncertainty in Motion Prediction with Variational Bayesian Mixture.** 15428-15437
- Daejun Kang, Dongsuk Kum, Sanmin Kim: **Continual Learning for Motion Prediction Model via Meta-Representation Learning and Optimal Memory Buffer Retention Strategy.** 15438-15448
- Xinshuo Weng, Boris Ivanovic, Yan Wang, Yue Wang, Marco Pavone: **PARA-Drive: Parallelized Architecture for Real-Time Autonomous Driving.** 15449-15458
- Jiawei Zhang, Chejian Xu, Bo Li: **ChatScene: Knowledge-Enabled Safety-Critical Scenario Generation for Autonomous Vehicles.** 15459-15469
- Lingjun Zhao, Jingyu Song, Katherine A. Skinner: **CRKD: Enhanced Camera-Radar Object Detection with Cross-Modality Knowledge Distillation.** 15470-15480
- Yue Hu, Junlong Peng, Sifei Liu, Junhao Ge, Si Liu, Siheng Chen: **Communication-Efficient Collaborative Perception via Information Filling with Codebook.** 15481-15490
- Geonho Bang, Kwangjin Choi, Jisong Kim, Dongsuk Kum, Jun Won Choi: **RadarDistill: Boosting Radar-Based Object Detection Performance via Knowledge Distillation from LiDAR Features.** 15491-15500
- Yancong Lin, Holger Caesar: **ICP-Flow: LiDAR Scene Flow Estimation with ICP.** 15501-15511
- Tianhao Zhao, Yongcan Chen, Yu Wu, Tianyang Liu, Bo Du, Peilun Xiao, Shi Qiu, Hongda Yang, Guozhen Li, Yi Yang, Yutian Lin: **Improving Bird's Eye View Semantic Segmentation by Task Decomposition.** 15512-15521
- Chen Min, Dawei Zhao, Liang Xiao, Jian Zhao, Xinli Xu, Zheng Zhu, Lei Jin, Jianshu Li, Yulan Guo, Junliang Xing, Liping Jing, Yiming Nie, Bin Dai: **DriveWorld: 4D Pre-Trained Scene Understanding via World Models for Autonomous Driving.** 15522-15533
- Chaohu Liu, Kun Yin, Haoyu Cao, Xinghua Jiang, Xin Li, Yinsong Liu, Deqiang Jiang, Xing Sun, Linli Xu: **HRVDA: High-Resolution Visual Document Assistant.** 15534-15545
- Xin Li, Yunfei Wu, Xinghua Jiang, Zhihao Guo, Mingming Gong, Haoyu Cao, Yinsong Liu, Deqiang Jiang, Xing Sun: **Enhancing Visual Document Understanding with Contrastive Learning in Large Visual-Language Models.** 15546-15555
- Yufan Chen, Jiaming Zhang, Kunyu Peng, Junwei Zheng, Ruiping Liu, Philip Torr, Rainer Stiefelhagen: **RoDLA: Benchmarking the Robustness of Document Layout Analysis Models.** 15556-15566
- Zhen Zhao, Jingqun Tang, Chunhui Lin, Binghong Wu, Can Huang, Hao Liu, Xin Tan, Zhizhong Zhang, Yuan Xie: **Multi-modal In-Context Learning Makes an Ego-evolving Scene Text Recognizer.** 15567-15576
- Changsheng Chen, Liangwei Lin, Yongqi Chen, Bin Li, Jishen Zeng, Jiwu Huang: **CMA: A Chromaticity Map Adapter for Robust Detection of Screen-Recapture Document Images.** 15577-15586
- Chen Duan, Pei Fu, Shan Guo, Qianyi Jiang, Xiaoming Wei: **ODM: A Text-Image Further Alignment Pre-training Approach for Scene Text Detection and Spotting.** 15587-15597
- Tsachi Blau, Sharon Fogel, Roi Ronen, Alona Golts, Shahar Tsiper, Elad Ben-Avraham, Aviad Aberdam, Roy Ganz, Ron Litman: **GRAM: Global Reasoning for Multi-Page VQA.** 15598-15607
- Mingxin Huang, Hongliang Li, Yuliang Liu, Xiang Bai, Lianwen Jin: **Bridging the Gap Between End-to-End and Two-Step Text Spotting.** 15608-15618
- Miao Rang, Zhenni Bi, Chuanjian Liu, Yunhe Wang, Kai Han: **An Empirical Study of Scaling Law for Scene Text Recognition.** 15619-15629
- Chuwei Luo, Yufan Shen, Zhaoqing Zhu, Qi Zheng, Zhi Yu, Cong Yao: **LayoutLLM: Layout Instruction Tuning with Large Language Models for Document Understanding.** 15630-15640



- Jianqiang Wan, Sibo Song, Wenwen Yu, Yuliang Liu, Wenqing Cheng, Fei Huang, Xiang Bai, Cong Yao, Zhibo Yang:
OMNIPARSER: A Unified Framework for Text Spotting, Key Information Extraction and Table Recognition. 15641-15653
- Jiaxin Zhang, Dezhi Peng, Chongyu Liu, Peirong Zhang, Lianwen Jin:
DocRes: A Generalist Model Toward Unifying Document Image Restoration Tasks. 15654-15664
- Min Liang, Jia-Wei Ma, Xiaobin Zhu, Jingyan Qin, Xu-Cheng Yin:
LayoutFormer: Hierarchical Text Detection Towards Scene Text Understanding. 15665-15674
- Yu Chen, Fei Gao, Yanguang Zhang, Maoying Qiao, Nannan Wang:
Generating Handwritten Mathematical Expressions From Symbol Graphs: An End-to-End Pipeline. 15675-15685
- Lingdong Kong, Youquan Liu, Lai Xing Ng, Benoit R. Cottereau, Wei Tsang Ooi:
OpenESS: Event-Based Semantic Scene Understanding with Open Vocabularies. 15686-15698
- Yangyang Guo, Guangzhi Wang, Mohan S. Kankanhalli:
PELA: Learning Parameter-Efficient Models with Low-Rank Approximation. 15699-15709
- Jianjian Cao, Peng Ye, Shengze Li, Chong Yu, Yansong Tang, Jiwen Lu, Tao Chen:
MADTP: Multimodal Alignment-Guided Dynamic Token Pruning for Accelerating Vision-Language Transformer. 15710-15719
- Roy Miles, Ismail Elezi, Jiankang Deng:
\$V_{\{k\}}D\$: Improving Knowledge Distillation Using Orthogonal Projections. 15720-15730
- Shangquan Sun, Wenqi Ren, Jingzhi Li, Rui Wang, Xiaochun Cao:
Logit Standardization in Knowledge Distillation. 15731-15740
- Sanghyeok Lee, Joonmyung Choi, Hyunwoo J. Kim:
Multi-Criteria Token Fusion with One-Step-Ahead Attention for Efficient Vision Transformers. 15741-15750
- Kai Han, Yunhe Wang, Jianyuan Guo, Enhua Wu:
ParameterNet: Parameters are All You Need for Large-Scale Visual Pretraining of Mobile Networks. 15751-15761
- Xinyin Ma, Gongfan Fang, Xinchao Wang:
DeepCache: Accelerating Diffusion Models for Free. 15762-15772
- Narges Norouzi, Svetlana Orlova, Daan de Geus, Gijs Dubbelman:
ALGM: Adaptive Local-then-Global Token Merging for Efficient Semantic Segmentation with Plain Vision Transformers. 15773-15782
- Wenxuan Huang, Yunhang Shen, Jiao Xie, Baochang Zhang, Gaoqi He, Ke Li, Xing Sun, Shaohui Lin:
A General and Efficient Training for Transformer via Token Expansion. 15783-15792
- Jianyang Gu, Saeed Vahidian, Vyacheslav Kungurtsev, Haonan Wang, Wei Jiang, Yang You, Yiran Chen:
Efficient Dataset Distillation via Minimax Diffusion. 15793-15803
- Niccolò Cavagnero, Gabriele Rosi, Claudia Cuttano, Francesca Pistilli, Marco Ciccone, Giuseppe Averta, Fabio Cermelli:
PEM: Prototype-Based Efficient MaskFormer for Image Segmentation. 15804-15813
- Jingxuan Xu, Wuyang Chen, Yao Zhao, Yunchao Wei:
Transferable and Principled Efficiency for Open-Vocabulary Segmentation. 15814-15824
- Hanxiao Zhang, Yifan Zhou, Guo-Hua Wang:
Dense Vision Transformer Compression with Few Samples. 15825-15834
- Chen Zhao, Shuming Liu, Karttikeya Mangalam, Guocheng Qian , Fatimah Zohra, Abdulmohsen Alghannam, Jitendra Malik, Bernard Ghanem :
Dr²Net: Dynamic Reversible Dual-Residual Networks for Memory-Efficient Finetuning. 15835-15844
- Jingyang Xiang, Siqi Li, Junhao Chen, Zhuangzhi Chen, Tianxin Huang, Linpeng Peng, Yong Liu:
MaxQ: Multi-Axis Query for N: m Sparsity Network. 15845-15854
- Chen Tang, Yuan Meng, Jiacheng Jiang, Shuzhao Xie, Rongwei Lu, Xinzhu Ma, Zhi Wang, Wenwu Zhu:
Retraining-free Model Quantization via One-Shot Weight-Coupling Learning. 15855-15865
- Jialin Li, Qiang Nie, Weifu Fu, Yuhuan Lin, Guangpin Tao, Yong Liu, Chengjie Wang:
LORS: Low-Rank Residual Structure for Parameter-Efficient Network Stacking. 15866-15876
- Ye Chen, Bingbing Ni, Jinfan Liu, Xiaoyang Huang, Xuanhong Chen:
Towards High-fidelity Artistic Image Vectorization via Texture-Encapsulated Shape Parameterization. 15877-15886
- Yonglong Tian, Lijie Fan, Kaifeng Chen, Dina Katabi, Dilip Krishnan, Phillip Isola:
Learning Vision from Models Rivals Learning Vision from Data. 15887-15898
- Yuzhang Shang, Dan Xu, Gaowen Liu, Ramana Rao Kompella, Yan Yan:
Efficient Multitask Dense Predictor via Binarization. 15899-15908
- Ao Wang, Hui Chen, Zijia Lin, Jungong Han, Guiguang Ding:
Rep ViT: Revisiting Mobile CNN From ViT Perspective. 15909-15920
- Yuzhang Shang, Gaowen Liu, Ramana Rao Kompella, Yan Yan:
Enhancing Post-Training Quantization Calibration Through Contrastive Learning. 15921-15930
- Yuan Zhang, Tao Huang, Jiaming Liu, Tao Jiang, Kuan Cheng, Shanghang Zhang:
FreeKD: Knowledge Distillation via Semantic Frequency Prompt. 15931-15940



- Chengtao Lv, Hong Chen, Jinyang Guo, Jinyang Guo, Jinyang Guo, Yifu Ding, Xianglong Liu:
PTQ4SAM: Post-Training Quantization for Segment Anything. 15941-15951
- Chuanguang Yang, Zhulin An, Libo Huang, Junyu Bi, Xinqiang Yu, Han Yang, Boyu Diao, Yongjun Xu:
CLIP-KD: An Empirical Study of CLIP Model Distillation. 15952-15962
- Pavan Kumar Anasosalu Vasu, Hadi Pouransari, Fartash Faghri, Raviteja Vemulapalli, Oncel Tuzel:
MobileCLIP: Fast Image-Text Models through Multi-Modal Reinforced Training. 15963-15974
- Shicai Wei, Chunbo Luo, Yang Luo:
Scale Decoupled Distillation. 15975-15983
- Nicolae-Catalin Ristea, Florinel-Alin Croitoru, Radu Tudor Ionescu, Marius Popescu, Fahad Shahbaz Khan, Mubarak Shah:
Self-Distilled Masked Auto-Encoders are Efficient Video Anomaly Detectors. 15984-15995
- Marina Neseem, Conor McCullough, Randy Hsin, Chas Leichner, Shan Li, In Suk Chong, Andrew G. Howard, Lukasz Lew, Sherief Reda, Ville-Mikko Rautio, Daniele Moro:
PikeLPN: Mitigating Overlooked Inefficiencies of Low-Precision Neural Networks. 15996-16005
- Fushuo Huo, Wenchao Xu, Jingcai Guo, Haozhao Wang, Song Guo:
C²KD: Bridging the Modality Gap for Cross-Modal Knowledge Distillation. 16006-16015
- Yu Wang, Xin Li, Shengzhao Weng, Gang Zhang, Haixiao Yue, Haocheng Feng, Junyu Han, Errui Ding:
KD-DETR: Knowledge Distillation for Detection Transformer with Consistent Distillation Points Sampling. 16016-16025
- Changyuan Wang, Ziwei Wang, Xiuwei Xu, Yansong Tang, Jie Zhou, Jiwen Lu:
Towards Accurate Post-Training Quantization for Diffusion Models. 16026-16035
- Qixuan Zheng, Ming Zhang, Hong Yan:
CURSOR: Scalable Mixed-Order Hypergraph Matching with CUR Decomposition. 16036-16045
- Andreas Bär, Neil Houlsby, Mostafa Dehghani, Manoj Kumar:
Frozen Feature Augmentation for Few-Shot Image Classification. 16046-16057
- Alireza Ganjdanesh, Shangqian Gao, Heng Huang:
Jointly Training and Pruning CNNs via Learnable Agent Guidance and Alignment. 16058-16069
- Hongjie Wang, Bhishma Dedhia, Niraj K. Jha:
Zero-TPrune: Zero-Shot Token Pruning Through Leveraging of the Attention Graph in Pre-Trained Transformers. 16070-16079
- Hongjie Wang, Difan Liu, Yan Kang, Yijun Li, Zhe Lin, Niraj K. Jha, Yuchen Liu:
Attention-Driven Training-Free Efficiency Enhancement of Diffusion Models. 16080-16089
- Shangqian Gao, Yanfu Zhang, Feihu Huang, Heng Huang:
BilevelPruning: Unified Dynamic and Static Channel Pruning for Convolutional Neural Networks. 16090-16100
- Wei Dong, Xing Zhang, Bihui Chen, Dawei Yan, Zhijun Lin, Qingsen Yan, Peng Wang, Yang Yang:
Low-Rank Rescaled Vision Transformer Fine-Tuning: A Residual Design Approach. 16101-16110
- Yunyang Xiong, Bala Varadarajan, Lemeng Wu, Xiaoyu Xiang, Fanyi Xiao, Chenchen Zhu, Xiaoliang Dai, Dilin Wang, Fei Sun, Forrest N. Iandola, Raghuraman Krishnamoorthi, Vikas Chandra:
EfficientSAM: Leveraged Masked Image Pretraining for Efficient Segment Anything. 16111-16121
- Lin Zhao, Tianchen Zhao, Zinan Lin, Xuefei Ning, Guohao Dai, Huazhong Yang, Yu Wang:
FlashEval: Towards Fast and Accurate Evaluation of Text-to-Image Diffusion Generative Models. 16122-16131
- Jaehyeon Moon, Dohyung Kim, Junyong Cheon, Bumsub Ham:
Instance-Aware Group Quantization for Vision Transformers. 16132-16141
- Leonardo Iurada, Marco Ciccone, Tatiana Tommasi:
Finding Lottery Tickets in Vision Models via Data-Driven Spectral Foresight Pruning. 16142-16151
- Kento Nishi, Junsik Kim, Wanhua Li, Hanspeter Pfister:
Joint-Task Regularization for Partially Labeled Multi-Task Learning. 16152-16162
- Xidong Wu, Shangqian Gao, Zeyu Zhang, Zhenzhen Li, Runxue Bao, Yanfu Zhang, Xiaoqian Wang, Heng Huang:
Auto-Train-Once: Controller Network Guided Automatic Network Pruning from Scratch. 16163-16173
- Yifu Ding, Weilun Feng, Chuyan Chen, Jinyang Guo, Xianglong Liu:
Reg-PTQ: Regression-specialized Post-training Quantization for Fully Quantized Object Detector. 16174-16184
- Matteo Farina, Massimiliano Mancini, Elia Cunegatti, Elia Cunegatti, Giovanni Iacca, Elisa Ricci:
MULTIFLOW: Shifting Towards Task-Agnostic Vision-Language Pruning. 16185-16195
- Ahmed Agiza, Marina Neseem, Sherief Reda:
MTLoRA: A Low-Rank Adaptation Approach for Efficient Multi-Task Learning. 16196-16205
- Fatih Ilhan, Gong Su, Selim Furkan Tekin, Tiansheng Huang, Sihao Hu, Ling Liu:
Resource- Efficient Transformer Pruning for Finetuning of Large Models. 16206-16215
- Minyoung Hwang, Luca Weihs, Chanwoo Park, Kimin Lee, Aniruddha Kembhavi, Kiana Ehsani:
Promptable Behaviors: Personalizing Multi-Objective Rewards from Human Preferences. 16216-16226
- Kiana Ehsani, Tanmay Gupta, Rose Hendrix, Jordi Salvador, Luca Weihs, Kuo-Hao Zeng, Kunal Pratap Singh, Yejin Kim, Winson Han, Alvaro Herrasti, Ranjay Krishna, Dustin Schwenk, Eli VanderBilt, Aniruddha Kembhavi:

**SPOC: Imitating Shortest Paths in Simulation Enables Effective Navigation and Manipulation in the Real World.**

16238-16250

- Zeyuan Yang, Jiageng Lin, Peihao Chen, Anoop Cherian, Tim K. Marks, Jonathan Le Roux, Chuang Gan:
RILA: Reflective and Imaginative Language Agent for Zero-Shot Semantic Audio-Visual Navigation. 16251-16261

- Yandan Yang, Baoxiong Jia, Peiyuan Zhi, Siyuan Huang:
PhyScene: Physically Interactable 3D Scene Synthesis for Embodied AI. 16262-16272

- Ram Ramrakhya, Aniruddha Kembhavi, Dhruv Batra, Zsolt Kira, Kuo-Hao Zeng, Luca Weihs:
Seeing the Unseen: Visual Common Sense for Semantic Placement. 16273-16283

- Yue Yang, Fan-Yun Sun, Luca Weihs, Eli VanderBilt, Alvaro Herrasti, Winson Han, Jiajun Wu, Nick Haber, Ranjay Krishna, Lingjie Liu, Chris Callison-Burch, Mark Yatskar, Aniruddha Kembhavi, Christopher Clark:
Holodeck: Language Guided Generation of 3D Embodied AI Environments. 16277-16287

- Yuhang Yang, Wei Zhai, Hongchen Luo, Yang Cao, Zheng-Jun Zha:
LEMON: Learning 3D Human-Object Interaction Relation from 2D Images. 16284-16295

- Ganlong Zhao, Guanbin Li, Weikai Chen, Yizhou Yu:
OVER-NAV: Elevating Iterative Vision-and-Language Navigation with Open-Vocabulary Detection and StructurEd Representation. 16296-16306

- Yiran Qin, Enshen Zhou, Qichang Liu, Zhenfei Yin, Lu Sheng, Ruimao Zhang, Yu Qiao, Jing Shao:
MP5: A Multi-modal Open-ended Embodied System in Minecraft via Active Perception. 16307-16316

- Rui Liu, Wenguan Wang, Yi Yang:
Volumetric Environment Representation for Vision-Language Navigation. 16317-16328

- Xiaohan Lei, Min Wang, Wengang Zhou, Li Li, Houqiang Li:
Instance-Aware Exploration-Verification-Exploitation for Instance ImageGoal Navigation. 16329-16339

- Ruihai Wu, Haoran Lu, Yiyan Wang, Yubo Wang, Hao Dong:
UniGarmentManip: A Unified Framework for Category-Level Garment Manipulation via Dense Visual Correspondence. 16340-16350

- Lei Fan, Mingfu Liang, Yunxuan Li, Gang Hua, Ying Wu:
Evidential Active Recognition: Intelligent and Prudent Open-World Embodied Perception. 16351-16361

- Zifan Wang, Junyu Chen, Ziqing Chen, Pengwei Xie, Rui Chen, Li Yi:
GenH2R: Learning Generalizable Human-to-Robot Handover via Scalable Simulation, Demonstration, and Imitation. 16362-16372

- Mukul Khanna, Ram Ramrakhya, Gunjan Chhablani, Sriram Yenamandra, Théophile Gervet, Matthew Chang, Zsolt Kira, Devendra Singh Chaplot, Dhruv Batra, Roozbeh Mottaghi:
GOAT-Bench: A Benchmark for Multi-Modal Lifelong Navigation. 16373-16383

- Mukul Khanna, Yongsen Mao, Hanxiao Jiang, Sanjay Haresh, Brennan Shacklett, Dhruv Batra, Alexander Clegg, Eric Undersander, Angel X. Chang, Manolis Savva:
Habitat Synthetic Scenes Dataset (HSSD-200): An Analysis of 3D Scene Scale and Realism Tradeoffs for ObjectGoal Navigation. 16384-16393

- Lei Fan, Jianxiong Zhou, Xiaoying Xing, Ying Wu:
Active Open-Vocabulary Recognition: Let Intelligent Moving Mitigate CLIP Limitations. 16394-16403

- Yichao Liang, Kevin Ellis, João Henriques:
Rapid Motor Adaptation for Robotic Manipulator Arms. 16404-16413

- Sixian Zhang, Xinyao Yu, Xinhang Song, Xiaohan Wang, Shuqiang Jiang:
Imagine Before Go: Self-Supervised Generative Map for Object Goal Navigation. 16414-16425

- Hao Li, Xue Yang, Zhaokai Wang, Xizhou Zhu, Jie Zhou, Yu Qiao, Xiaogang Wang, Hongsheng Li, Lewei Lu, Jifeng Dai:
Auto MC-Reward: Automated Dense Reward Design with Large Language Models for Minecraft. 16426-16435

- Xiao Chen, Quanyi Li, Tai Wang, Tianfan Xue, Jiangmiao Pang:
GenNBV: Generalizable Next-Best-View Policy for Active 3D Reconstruction. 16436-16445

- Xiaohan Wang, Yuehu Liu, Xinhang Song, Yuyi Liu, Sixian Zhang, Shuqiang Jiang:
An Interactive Navigation Method with Effect-oriented Affordance. 16446-16456

- Zhixuan Liang, Yao Mu, Hengbo Ma, Masayoshi Tomizuka, Mingyu Ding, Ping Luo:
SkillDiffuser: Interpretable Hierarchical Planning via Skill Abstractions in Diffusion-Based Task Execution. 16467-16476

- Ziwei Zhao, Yuchen Wang, Chuhua Wang:
Fusing Personal and Environmental Cues for Identification and Segmentation of First-Person Camera Wearers in Third-Person Views. 16477-16487

- Arjun Majumdar, Anurag Ajay, Xiaohan Zhang, Pranav Putta, Sriram Yenamandra, Mikael Henaff, Sneha Silwal, Paul McVay, Oleksandr Maksymets, Sergio Arnaud, Karmesh Yadav, Qiyang Li, Ben Newman, Mohit Sharma, Vincent-Pierre Berges, Shiqi Zhang, Pulkit Agrawal, Yonatan Bisk, Dhruv Batra, Mrinal Kalakrishnan, Franziska Meier, Chris Paxton, Alexander Sax, Aravind Rajeswaran:
OpenEQA: Embodied Question Answering in the Era of Foundation Models. 16488-16498

- Jaehyun Song, Minjong Yoo, Honguk Woo:
Model Adaptation for Time Constrained Embodied Control. 16499-16508



- Subhadeep Koley, Ayan Kumar Bhunia, Aneeshan Sain, Pinaki Nath Chowdhury, Tao Xiang, Yi-Zhe Song: **You'll Never Walk Alone: A Sketch and Text Duet for Fine-Grained Image Retrieval.** 16509-16519
- Jiabao Wang, Yuming Chen, Zhaohui Zheng, Xiang Li, Ming-Ming Cheng, Qibin Hou: **CrossKD: Cross-Head Knowledge Distillation for Object Detection.** 16520-16530
- Tz-Ying Wu, Chih-Hui Ho, Nuno Vasconcelos: **ProTeCt: Prompt Tuning for Taxonomic Open Set Classification.** 16531-16540
- Mikhail Kennerley, Jian-Gang Wang, Bharadwaj Veeravalli, Robby T. Tan: **CAT: Exploiting Inter-Class Dynamics for Domain Adaptive Object Detection.** 16541-16550
- Jiamian Wang, Pichao Wang, Guohao Sun, Dongfang Liu, Sohail A. Dianat, Raghubeer Rao, Majid Rabbani, Zhiqiang Tao: **Text Is MASS: Modeling as Stochastic Embedding for Text-Video Retrieval.** 16551-16560
- Zhuoling Li, Xiaogang Xu, Ser-Nam Lim, Hengshuang Zhao: **UniMODE: Unified Monocular 3D Object Detection.** 16561-16570
- Zehong Ma , Shiliang Zhang, Longhui Wei, Qi Tian: **OVMR: Open-Vocabulary Recognition with Multi-Modal References.** 16571-16581
- Yong-Lu Li, Xiaoqian Wu, Xinpeng Liu, Zehao Wang, Yiming Dou, Yikun Ji, Junyi Zhang, Yixing Li, Xudong Lu, Jingru Tan, Cewu Lu: **From Isolated Islands to Pangea: Unifying Semantic Space for Human Action Understanding.** 16582-16592
- Jang Hyun Cho, Philipp Krähenbühl: **Language-Conditioned Detection Transformer.** 16593-16603
- Kunlun Xu, Xu Zou, Yuxin Peng, Jiahuan Zhou: **Distribution-Aware Knowledge Prototyping for Non-Exemplar Lifelong Person Re-Identification.** 16604-16613
- Zhenyu Cui, Jiahuan Zhou, Xun Wang, Manyu Zhu, Yuxin Peng: **Learning Continual Compatible Representation for Re-indexing Free Lifelong Person Re-identification.** 16614-16623
- Dejie Yang, Yang Liu: **Active Object Detection with Knowledge Aggregation and Distillation from Large Models.** 16624-16633
- Mingxuan Liu, Tyler L. Hayes, Elisa Ricci, Gabriela Csurka, Riccardo Volpi: **SHiNe: Semantic Hierarchy Nexus for Open-Vocabulary Object Detection.** 16634-16644
- Kaiyu Yue, Bor-Chun Chen, Jonas Geiping, Hengduo Li, Tom Goldstein, Ser-Nam Lim: **Object Recognition as Next Token Prediction.** 16645-16656
- Ting Lei, Shaofeng Yin, Yang Liu: **Exploring the Potential of Large Foundation Models for Open-Vocabulary HOI Detection.** 16657-16667
- Jiangpeng He: **Gradient Reweighting: Towards Imbalanced Class-Incremental Learning.** 16668-16677
- Jiaming Li, Jiacheng Zhang, Jichang Li, Ge Li, Si Liu, Liang Lin, Guanbin Li: **Learning Background Prompts to Discover Implicit Knowledge for Open Vocabulary Object Detection.** 16678-16687
- Xianpeng Liu, Ce Zheng, Ming Qian, Nan Xue, Chen Chen, Zhebin Zhang, Chen Li, Tianfu Wu: **Multi-View Attentive Contextualization for Multi-View 3D Object Detection.** 16688-16698
- Ximiao Zhang, Min Xu, Xiuzhuang Zhou: **RealNet: A Feature Selection Network with Realistic Synthetic Anomaly for Anomaly Detection.** 16699-16708
- Shitong Shao, Zeyuan Yin, Muxin Zhou, Xindong Zhang, Zhiqiang Shen: **Generalized Large-Scale Data Condensation via Various Backbone and Statistical Matching.** 16709-16718
- Guopeng Li, Ming Qian, Gui-Song Xia: **Unleashing Unlabeled Data: A Paradigm for Cross-View Geo-Localization.** 16719-16729
- Junwei Luo, Xue Yang, Yi Yu, Qingyun Li, Junchi Yan, Yansheng Li: **PointOBB: Learning Oriented Object Detection via Single Point Supervision.** 16730-16740
- Xiaowei Zhao, Xianglong Liu, Duorui Wang, Yajun Gao, Zhide Liu: **Scene-adaptive and Region-aware Multi-modal Prompt for Open Vocabulary Object Detection.** 16741-16750
- Wenqiao Zhang, Zheqi Lv: **Revisiting the Domain Shift and Sample Uncertainty in Multi-source Active Domain Transfer.** 16751-16761
- Fanjie Kong, Yanbei Chen, Jiarui Cai, Davide Modolo: **Hyperbolic Learning with Synthetic Captions for Open-World Detection.** 16762-16771
- Feng Lu, Xiangyuan Lan, Lijun Zhang, Dongmei Jiang, Yaowei Wang, Chun Yuan: **CricaVPR: Cross-Image Correlation-Aware Representation Learning for Visual Place Recognition.** 16772-16782
- Yi Yu, Xue Yang, Qingyun Li, Feipeng Da, Jifeng Dai, Yu Qiao, Junchi Yan: **Point2RBox: Combine Knowledge from Synthetic Visual Patterns for End-to-End Oriented Object Detection with Single Point Supervision.** 16783-16793
- Yansong Peng, Hebei Li, Yueyi Zhang, Xiaoyan Sun, Feng Wu: **Scene Adaptive Sparse Transformer for Event-based Object Detection.** 16794-16804



- Young Kyun Jang, Donghyun Kim, Zihang Meng, Dat Huynh, Ser-Nam Lim:
Visual Delta Generator with Large Multi-Modal Models for Semi-Supervised Composed Image Retrieval. 16805-16814
- Li Lin, Xianan He, Yan Ju, Xin Wang, Feng Ding, Shu Hu:
Preserving Fairness Generalization in Deepfake Detection. 16815-16825
- Subhadeep Koley, Ayan Kumar Bhunia, Aneeshan Sain, Pinaki Nath Chowdhury, Tao Xiang, Yi-Zhe Song:
Text-to-Image Diffusion Models are Great Sketch-Photo Matchmakers. 16826-16837
- Zhi-Fan Wu, Chaojie Mao, Xue Wang, Jianwen Jiang, Yiliang Lv, Rong Jin:
Structured Model Probing: Empowering Efficient Transfer Learning by Structured Regularization. 16838-16847
- Xiaofan Li, Zhizhong Zhang, Xin Tan, Chengwei Chen, Yanyun Qu, Yuan Xie, Lizhuang Ma:
PromptAD: Learning Prompts with only Normal Samples for Few-Shot Anomaly Detection. 16848-16858
- Subhadeep Koley, Ayan Kumar Bhunia, Aneeshan Sain, Pinaki Nath Chowdhury, Tao Xiang, Yi-Zhe Song:
How to Handle Sketch-Abstraction in Sketch-Based Image Retrieval? 16859-16869
- Bin Yang, Jun Chen, Mang Ye:
Shallow-Deep Collaborative Learning for Unsupervised Visible-Infrared Person Re-Identification. 16870-16879
- Xinzi Cao, Xiawu Zheng, Guanhong Wang, Weijiang Yu, Yunhang Shen, Ke Li, Yutong Lu, Yonghong Tian:
Solving the Catastrophic Forgetting Problem in Generalized Category Discovery. 16880-16889
- Shijie Ma, Fei Zhu, Zhun Zhong, Xu-Yao Zhang, Cheng-Lin Liu:
Active Generalized Category Discovery. 16890-16900
- Tianheng Cheng, Lin Song, Yixiao Ge, Wenyu Liu, Xinggang Wang, Ying Shan:
YOLO-World: Real-Time Open-Vocabulary Object Detection. 16901-16911
- Zi-Kai Xiao, Guo-Ye Yang, Xue Yang, Tai-Jiang Mu, Junchi Yan, Shi-Min Hui:
Theoretically Achieving Continuous Representation of Oriented Bounding Boxes. 16912-16922
- Jiacheng Zhang, Jiaming Li, Xiangru Lin, Wei Zhang, Xiao Tan, Junyu Han, Errui Ding, Jingdong Wang, Guanbin Li:
Decoupled Pseudo-Labeling for Semi-Supervised Monocular 3D Object Detection. 16923-16932
- Ziyi Wu, Mathias Gehrig, Qing Lyu, Xudong Liu, Igor Gilitschenski:
LEOD: Label-Efficient Object Detection for Event Cameras. 16933-16942
- Kunyang Zhou:
Lane2Seq: Towards Unified Lane Detection via Sequence Generation. 16944-16953
- Jie Yang, Bingliang Li, Ailing Zeng, Lei Zhang, Ruimao Zhang:
Open-World Human-Object Interaction Detection via Multi-Modal Prompts. 16954-16964
- Yian Zhao, Wenyu Lv, Shangliang Xu, Jinman Wei, Guanzhong Wang, Qingqing Dang, Yi Liu, Jie Chen:
DETRs Beat YOLOs on Real-time Object Detection. 16965-16974
- Jaewoo Jeong, Daehee Park, Kuk-Jin Yoon:
Multi-Agent Long-Term 3D Human Pose Forecasting via Interaction-Aware Trajectory Conditioning. 16975-16984
- Heng Zhang, Qiuyu Zhao, Linyu Zheng, Hao Zeng, Zhiwei Ge, Tianhao Li, Sulong Xu:
Exploring Region-Word Alignment in Built-in Detector for Open-Vocabulary Object Detection. 16975-16984
- Siyang Dai, Jun Liu, Ngai-Man Cheung:
Referring Expression Counting. 16985-16995
- Wenshuai Xu, Zhenghui Hu, Yu Lu, Jinzhou Meng, Qingjie Liu, Yunhong Wang:
ActiveDC: Distribution Calibration for Active Finetuning. 16996-17005
- Yunpeng Luo, Junlong Du, Ke Yan, Shouhong Ding:
LaRE²: Latent Reconstruction Error Based Method for Diffusion-Generated Image Detection. 17006-17015
- Fan Zhang, Xian-Sheng Hua, Chong Chen, Xiao Luo:
Fine-grained Prototypical Voting with Heterogeneous Mixup for Semi-supervised 2D-3D Cross-modal Retrieval. 17016-17026
- Chuyang Zhao, Yifan Sun, Wenhao Wang, Qiang Chen, Errui Ding, Yi Yang, Jingdong Wang:
MS-DETR: Efficient DETR Training with Mixed Supervision. 17027-17036
- Yun Li, Zhe Liu, Hang Chen, Lina Yao:
Context-Based and Diversity-Driven Specificity in Compositional Zero-Shot Learning. 17037-17046
- Yixuan Sun, Zhangyue Yin, Haibo Wang, Yan Wang, Xipeng Qiu, Weifeng Ge, Wenqiang Zhang:
Pixel-Level Semantic Correspondence Through Layout-Aware Representation Learning and Multi-Scale Matching Integration. 17047-17056
- Wenxiao Deng, Wenbin Li, Tianyu Ding, Lei Wang, Hongguang Zhang, Kuihua Huang, Jing Huo, Yang Gao:
Exploiting Inter-sample and Inter-feature Relations in Dataset Distillation. 17057-17066
- Zhizhong Huang, Mingliang Dai, Yi Zhang, Junping Zhang, Hongming Shan:
Point, Segment and Count: A Generalized Framework for Object Counting. 17067-17076
- Rohan Sarkar, Avinash C. Kak:
Dual Pose-invariant Embeddings: Learning Category and Object-specific Discriminative Representations for Recognition and Retrieval. 17077-17085



- Ziheng Chen, Yue Song, Gaowen Liu, Ramana Rao Kompella, Xiao-Jun Wu, Nicu Sebe:
Riemannian Multinomial Logistics Regression for SPD Neural Networks. 17086-17096
- Qin Zhang, Dongsheng An, Tianjun Xiao, Tong He, Qingming Tang, Ying Nian Wu, Joseph Tighe, Yifan Xing:
Learning for Transductive Threshold Calibration in Open-World Recognition. 17097-17106
- Michal Shlapentokh-Rothman, Ansel Blume, Yao Xiao, Yuqun Wu, Sethuraman TV, Heyi Tao, Jae Yong Lee, Wilfredo Torres, Yu-Xiong Wang, Derek Hoiem:
Region-Based Representations Revisited. 17107-17116
- Pingping Zhang, Yuhao Wang, Yang Liu, Zhengzheng Tu, Huchuan Lu:
Magic Tokens: Select Diverse Tokens for Multi-modal Object Re-Identification. 17117-17126
- Wentao Tan, Changxing Ding, Jiayu Jiang, Fei Wang, Yibing Zhan, Dapeng Tao:
Harnessing the Power of MLLMs for Transferable Text-to-Image Person ReID. 17127-17137
- Kaibin Tian, Ruixiang Zhao, Zijie Xin, Bangxiang Lan, Xirong Li:
Holistic Features are Almost Sufficient for Text-to-Video Retrieval. 17138-17147
- Feng Xue, Zi He, Yuan Zhang, Chuanlong Xie, Zhenguo Li, Falong Tan:
Enhancing the Power of OOD Detection via Sample-Aware Model Selection. 17148-17157
- Yuqi Wang, Yuntao Chen, Xingyu Liao, Lue Fan, Zhaoxiang Zhang:
PanoOcc: Unified Occupancy Representation for Camera-based 3D Panoptic Segmentation. 17158-17168
- Ziyang Luo, Nian Liu, Wangbo Zhao, Xuguang Yang, Dingwen Zhang, Deng-Ping Fan, Fahad Khan, Junwei Han:
VSCODE: General Visual Salient and Camouflaged Object Detection with 2D Prompt Learning. 17169-17180
- Yi Xie, Yihong Lin, Wenjie Cai, Xuemiao Xu, Huaidong Zhang, Yong Du, Shengfeng He:
D3still: Decoupled Differential Distillation for Asymmetric Image Retrieval. 17181-17190
- Yimeng Fan, Wei Zhang, Changsong Liu, Mingyang Li, Wenrui Lu:
SFOD: Spiking Fusion Object Detector. 17191-17200
- Liqiong Wang, Jinyu Yang, Yanfu Zhang, Fangyi Wang, Feng Zheng:
Depth-Aware Concealed Crop Detection in Dense Agricultural Scenes. 17201-17211
- Hyeyonjun Lee, Sehyun Hwang, Suha Kwak:
Extreme Point Supervised Instance Segmentation. 17212-17222
- Zhicai Wang, Longhui Wei, Tan Wang, Heyu Chen, Yanbin Hao, Xiang Wang, Xiangnan He, Qi Tian:
Enhance Image Classification via Inter-Class Image Mixup with Diffusion Model. 17223-17233
- Alex Costanzino, Pierluigi Zama Ramirez, Giuseppe Lisanti, Luigi Di Stefano:
Multimodal Industrial Anomaly Detection by Crossmodal Feature Mapping. 17234-17243
- Alex Warren, Ke Xu, Jiaying Lin, Gary K. L. Tam, Rynson W. H. Lau, Rynson W. H. Lau:
Effective Video Mirror Detection with Inconsistent Motion Cues. 17244-17252
- Can Xu, Yuehui Han, Rui Xu, Le Hui, Jin Xie, Jian Yang:
Multi-Attribute Interactions Matter for 3D Visual Grounding. 17253-17262
- Ankan Bhunia, Changjian Li, Hakan Bilen:
Looking 3D: Anomaly Detection with 2D-3D Alignment. 17263-17272
- Zhen-Duo Chen, Li-Jun Zhao, Zi-Chao Zhang, Xin Luo, Xin-Shun Xu:
Characteristics Matching Based Hash Codes Generation for Efficient Fine-Grained Image Retrieval. 17273-17281
- Yulu Gao, Yifan Sun, Xudong Ding, Chuyang Zhao, Si Liu:
EASE-DETR: Easing the Competition among Object Queries. 17282-17291
- Kaipeng Fang, Jingkuan Song, Lianli Gao, Pengpeng Zeng, Zhi-Qi Cheng, Xiye Li, Heng Tao Shen:
ProS: Prompting-to-Simulate Generalized Knowledge for Universal Cross-Domain Retrieval. 17292-17301
- Zhicheng Sun, Jinghan Li, Yadong Mu:
Exploring Orthogonality in Open World Object Detection. 17302-17312
- Mathilde Caron, Ahmet Iscen, Alireza Fathi, Cordelia Schmid:
A Generative Approach for Wikipedia-Scale Visual Entity Recognition. 17313-17322
- Ke Li, Di Wang, Zhangyuan Hu, Wenxuan Zhu, Shaofeng Li, Quan Wang:
Unleashing Channel Potential: Space-Frequency Selection Convolution for SAR Object Detection. 17323-17332
- Mohammad Saeed Ebrahimi Saadabadi, Ali Dabouei, Sahar Rahimi Malakshan, Nasser M. Nasrabadi:
Hyperspherical Classification with Dynamic Label-to-Prototype Assignment. 17333-17342
- Zexian Yang, Dayan Wu, Chenming Wu, Zheng Lin, Jingzi Gu, Weiping Wang:
A Pedestrian is Worth One Prompt: Towards Language Guidance Person Re-Identification. 17343-17353
- Zihua Liu, Hiroki Sakuma, Masatoshi Okutomi:
VSRD: Instance-Aware Volumetric Silhouette Rendering for Weakly Supervised 3D Object Detection. 17354-17363
- Hyeongjun Kwon, Jinhyun Jang, Jin Kim, Kwonyoung Kim, Kwanghoon Sohn, Kwanghoon Sohn:
Improving Visual Recognition with Hyperbolical Visual Hierarchy Mapping. 17364-17374
- Chull Hwan Song, Jooyoung Yoon, Taebaek Hwang, Shunghyun Choi, Yeong Hyeon Gu, Yannis Avrithis:
On Train-Test Class Overlap and Detection for Image Retrieval. 17375-17384



- Menghao Zhang, Jingyu Wang, Qi Qi, Haifeng Sun, Zirui Zhuang, Pengfei Ren, Rui long Ma, Jianxin Liao:
Multi-Scale Video Anomaly Detection by Multi-Grained Spatio-Temporal Representation Learning. 17385-17394
- Dat Nguyen, Nesryne Mejri, Inder Pal Singh, Polina Kuleshova, Marcella Astrid, Anis Kacem, Enjie Ghorbel, Djamila Aouada:
LAA-Net: Localized Artifact Attention Network for Quality-Agnostic and Generalizable Deepfake Detection. 17395-17405
- Hang Xu, Xinyuan Liu, Haonan Xu, Yike Ma, Zunjie Zhu, Chenggang Yan, Feng Dai:
Rethinking Boundary Discontinuity Problem for Oriented Object Detection. 17406-17415
- Jinjing Zhao, Fangyun Wei, Chang Xu:
Hybrid Proposal Refiner: Revisiting DETR Series from the Faster R-CNN Perspective. 17416-17426
- Jooyeon Kim, Eulrang Cho, Sehyung Kim, Hyunwoo J. Kim:
Retrieval-Augmented Open-Vocabulary Object Detection. 17427-17436
- Wenxuan Guo, Zhiyu Pan, Yingping Liang, Ziheng Xi, Zhicheng Zhong, Jianjiang Feng, Jie Zhou:
LiDAR-Based Person Re-Identification. 17437-17447
- Xu Zheng, Lin Wang:
EventDance: Unsupervised Source-Free Cross-Modal Adaptation for Event-Based Object Recognition. 17448-17458
- He Li, Mang Ye, Ming Zhang, Bo Du:
All in One Framework for Multimodal Re-Identification in the Wild. 17459-17469
- Prakhar Kaushik, Adam Kortylewski, Alan L. Yuille:
A Bayesian Approach to OOD Robustness in Image Classification. 17459-17469
- Bruce A. Maxwell, Sumegha Singhania, Avnish Patel, Rahul Kumar, Heather Fryling, Sihan Li, Haonan Sun, Ping He, Zewen Li:
Logarithmic Lenses: Exploring Log RGB Data for Image Classification. 17470-17479
- Yichen Bai, Zongbo Han, Bing Cao, Xiaoheng Jiang, Qinghua Hu, Changqing Zhang:
ID-like Prompt Learning for Few-Shot Out-of-Distribution Detection. 17480-17489
- Qiankun Liu, Rui Liu, Bolun Zheng, Hongkui Wang, Ying Fu:
Infrared Small Target Detection with Scale and Location Sensitivity. 17490-17499
- Yuting Li, Yingyi Chen, Xuanlong Yu, Dexiong Chen, Xi Shen:
SURE: SUrvey REcipes for Building Reliable and Robust Deep Networks. 17500-17510
- Huimin Li, Zhentao Chen, Yunhao Xu, Junlin Hu:
Hyperbolic Anomaly Detection. 17511-17520
- Weizhen He, Yiheng Deng, Shixiang Tang, Qihao Chen, Qingsong Xie, Yizhou Wang, Lei Bai, Feng Zhu, Rui Zhao, Wanli Ouyang, Donglian Qi, Yunfeng Yan:
Instruct-ReID: A Multi-Purpose Person Re-Identification Task with Instructions. 17521-17531
- Yiyu Chen, Zheyi Fan, Zhaoru Chen, Yixuan Zhu:
CA-Jaccard: Camera-aware Jaccard Distance for Person Re-identification. 17532-17541
- Oindrila Saha, Grant Van Horn, Subhransu Maji:
Improved Zero-Shot Classification by Adapting VLMs with Text Descriptions. 17542-17552
- Imad Eddine Toubal, Aditya Avinash, Neil Gordon Alldrin, Jan Dlabal, Wenlei Zhou, Enming Luo, Otilia Stretcu, Hao Xiong, Chun-Ta Lu, Howard Zhou, Ranjay Krishna, Ariel Fuxman, Tom Duerig:
Modeling Collaborator: Enabling Subjective Vision Classification with Minimal Human Effort via LLM Tool-Use. 17553-17563
- Emmanuel Onzon, Maximilian Bömer, Fahim Mannan, Felix Heide:
Neural Exposure Fusion for High-Dynamic Range Object Detection. 17564-17573
- Xiuquan Hou, Meiqin Liu, Senlin Zhang, Ping Wei, Badong Chen:
Salience DETR: Enhancing Detection Transformer with Hierarchical Salience Filtering Refinement. 17574-17583
- Tianqi Li, Guansong Pang, Xiao Bai, Wenjun Miao, Jin Zheng:
Learning Transferable Negative Prompts for Out-of-Distribution Detection. 17584-17594
- Guohao Peng, Heshan Li, Yangyang Zhao, Jun Zhang, Zhenyu Wu, Pengyu Zheng, Danwei Wang:
TransLoc4D: Transformer-Based 4D Radar Place Recognition. 17595-17605
- Deng Li, Aming Wu, Yaowei Wang, Yahong Han:
Prompt-Driven Dynamic Object-Centric Learning for Single Domain Generalization. 17606-17615
- Jiawen Zhu, Choubo Ding, Yu Tian, Guansong Pang:
Anomaly Heterogeneity Learning for Open-Set Supervised Anomaly Detection. 17616-17626
- Cheng-Yao Hong, Yen-Chi Hsu, Tyng-Luh Liu:
Contrastive Learning for DeepFake Classification and Localization via Multi-Label Ranking. 17627-17637
- Binrui Shen, Qiang Niu, Shengxin Zhu:
Adaptive Softassign via Hadamard-Equipped Sinkhorn. 17638-17647
- Feiran Hu, Chen-Lin Zhang, Jiangliang Guo, Xiu-Shen Wei, Lin Zhao, Anqi Xu, Lingyan Gao:
An Asymmetric Augmented Self-Supervised Learning Method for Unsupervised Fine-Grained Image Hashing. 17648-17657



- Sergio Izquierdo, Javier Civera:
Optimal Transport Aggregation for Visual Place Recognition. 17658-17668
- Martijn Oldenhof, Edward De Brouwer, Adam Arany, Yves Moreau:
Atom-Level Optical Chemical Structure Recognition with Limited Supervision. 17669-17678
- Yu Liu, Yaqi Cai, Qi Jia, Binglin Qiu, Weimin Wang, Nan Pu:
Novel Class Discovery for Ultra-Fine-Grained Visual Categorization. 17679-17688
- Yan Huang, Zhang Zhang, Qiang Wu, Yi Zhong, Liang Wang:
Attribute-Guided Pedestrian Retrieval: Bridging Person Re-ID with Internal Attribute Variability. 17689-17699
- Yuchen Yang, Likai Wang, Erkun Yang, Cheng Deng:
Robust Noisy Correspondence Learning with Equivariant Similarity Consistency. 17700-17709
- Ziteng Gao, Zhan Tong, Kevin Qinghong Lin, Joya Chen, Mike Zheng Shou:
Bootstrapping SparseFormers from Vision Foundation Models. 17710-17721
- Jae Hyeon Park, Gyoomin Lee, Seunggi Park, Sung In Cho:
Not All Classes Stand on Same Embeddings: Calibrating a Semantic Distance with Metric Tensor. 17722-17731
- Muhammad Sohail Danish, Muhammad Haris Khan, Muhammad Akhtar Munir, M. Saquib Sarfraz, Mohsen Ali:
Improving Single Domain-Generalized Object Detection: A Focus on Diversification and Alignment. 17732-17742
- Mubariz Zaffar, Liangliang Nan, Julian F. P. Kooij:
On the Estimation of Image-Matching Uncertainty in Visual Place Recognition. 17743-17753
- Aimira Baitieva, David Hurych, Victor Besnier, Olivier Bernard:
Supervised Anomaly Detection for Complex Industrial Images. 17754-17762
- Puru Vaish, Shunxin Wang, Nicola Strisciuglio:
Fourier-Basis Functions to Bridge Augmentation Gap: Rethinking Frequency Augmentation in Image Classification. 17763-17772
- Dai Shi:
TransNeXt: Robust Foveal Visual Perception for Vision Transformers. 17773-17783
- Chenhongyi Yang, Lichao Huang, Elliot J. Crowley:
Plug and Play Active Learning for Object Detection. 17784-17793
- Amar Ali-bey, Brahim Chaib-draa, Philippe Giguère:
BoQ: A Place is Worth a Bag of Learnable Queries. 17794-17803
- Nico Lang, Vésteinn Snæbjarnarson, Elijah Cole, Oisin Mac Aodha, Christian Igel, Serge J. Belongie:
From Coarse to Fine-Grained Open-Set Recognition. 17804-17814
- Eastman Z. Y. Wu, Yali Li, Yuan Wang, Shengjin Wang:
Exploring Pose-Aware Human-Object Interaction via Hybrid Learning. 17815-17825
- Jiawen Zhu, Guansong Pang:
Toward Generalist Anomaly Detection via In-Context Residual Learning with Few-Shot Sample Prompts. 17826-17836
- Guillaume Bono, Hervé Poirier, Leonid Antsfeld, Gianluca Monaci, Boris Chidlovskii, Christian Wolf:
Learning to Navigate Efficiently and Precisely in Real Environments. 17837-17846
- Pierre Marza, Laëtitia Matignon, Olivier Simonin, Christian Wolf:
Task-Conditioned Adaptation of Visual Features in Multi-Task Policy Learning. 17847-17856
- Yifei Zhang, Hao Zhao, Hongyang Li, Siheng Chen:
FastMAC: Stochastic Spectral Sampling of Correspondence Graph. 17857-17867
- Bowen Wen, Wei Yang, Jan Kautz, Stan Birchfield:
FoundationPose: Unified 6D Pose Estimation and Tracking of Novel Objects. 17868-17879
- Jiayi Liu, Hou In Ivan Tam, Ali Mahdavi-Amiri, Manolis Savva:
CAGE: Controllable Articulation GEneration. 17880-17889
- Inhwan Bae, Young-Jae Park, Hae-Gon Jeon:
SingularTrajectory: Universal Trajectory Predictor Using Diffusion Model. 17890-17901
- Vuong Dinh An , Minh Nhat Vu, Baoru Huang, Nghia Nguyen, Hieu Le, Thieu Vo, Anh Nguyen:
Language-driven Grasp Detection. 17902-17912
- Hongxin Li, Zeyu Wang, Xu Yang, Yuran Yang, Shuqi Mei, Zhaoxiang Zhang:
MemoNav: Working Memory Model for Visual Navigation. 17913-17922
- Van Nguyen Nguyen, Thibault Groueix, Georgy Poniatkin, Yinlin Hu, Renaud Marlet, Mathieu Salzmann, Vincent Lepetit:
NOPE: Novel Object Pose Estimation from a Single Image. 17923-17932
- Guo-Hao Xu, Yi-Lin Wei, Dian Zheng, Xiao-Ming Wu, Wei-Shi Zheng:
Dexterous Grasp Transformer. 17933-17942
- Gengyu Zhang, Hao Tang, Yan Yan:
Versatile Navigation Under Partial Observability via Value-Guided Diffusion Policy. 17943-17951
- Jun Wang, Yuzhe Qin, Kaiming Kuang, Yigit Korkmaz, Akhilan Gurumoorthy, Hao Su, Xiaolong Wang:
CyberDemo: Augmenting Simulated Human Demonstration for Real-World Dexterous Manipulation. 17952-17963



- Yunfei Fan, Tianyu Zhao, Guidong Wang:
SchurVINS: Schur Complement-Based Lightweight Visual Inertial Navigation System. 17964-17973
- Takeru Oba, Matthew R. Walter, Norimichi Ukita:
READ: Retrieval-Enhanced Asymmetric Diffusion for Motion Planning. 17974-17984
- Yichen Zhu, Zhicai Ou, Xiaofeng Mou, Jian Tang:
Retrieval-Augmented Embodied Agents. 17985-17995
- Rui Song, Chenwei Liang, Hu Cao, Zhiran Yan, Walter Zimmer, Markus Gross, Andreas Festag, Alois Knoll:
Collaborative Semantic Occupancy Prediction with Hybrid Feature Fusion in Connected Automated Vehicles. 17996-18006
- Hyunwoo Ryu, Jiwoo Kim, Hyunseok An, Junwoo Chang, Joohwan Seo, Taehan Kim, Yubin Kim, Chaewon Hwang, Jongeun Choi, Roberto Horowitz:
Diffusion-EDFs: Bi-Equivariant Denoising Generative Modeling on SE(3) for Visual Robotic Manipulation. 18007-18018
- Youqi Pan, Wugen Zhou, Yingdian Cao, Hongbin Zha:
Adaptive VIO: Deep Visual-Inertial Odometry with Online Continual Learning. 18019-18028
- Changan Chen, Rui Wang, Christoph Vogel, Marc Pollefeys:
F³Loc: Fusion and Filtering for Floorplan Localization. 18029-18038
- Hidenobu Matsuki, Riku Murai, Paul H. J. Kelly, Andrew J. Davison:
Gaussian Splatting SLAM. 18039-18048
- Shizhe Chen, Ricardo Garcia, Ivan Laptev, Cordelia Schmid:
SUGAR : Pre-training 3D Visual Representations for Robotics. 18049-18060
- Xiaoqi Li, Mingxu Zhang, Yiran Geng, Haoran Geng, Yuxing Long, Yan Shen, Renrui Zhang, Jiaming Liu, Hao Dong:
ManipLLM: Embodied Multimodal Large Language Model for Object-Centric Robotic Manipulation. 18061-18070
- Jaime Corsetti, Davide Boscaini, Changjae Oh, Andrea Cavallaro, Fabio Poiesi:
Open-vocabulary object 6D pose estimation. 18071-18080
- Xiao Ma, Sumit Patidar, Iain Haughton, Stephen James:
Hierarchical Diffusion Policy for Kinematics-Aware Multi-Task Robotic Manipulation. 18081-18090
- Zhihao Cao, Zidong Wang, Siwen Xie, Anji Liu, Lifeng Fan:
Smart Help: Strategic Opponent Modeling for Proactive and Adaptive Robot Assistance in Households. 18091-18101
- Haoxiang Ma, Modi Shi, Boyang Gao, Di Huang:
Generalizing 6-DoF Grasp Detection via Domain Prior Knowledge. 18102-18111
- Hongwei Ren, Jiadong Zhu, Yue Zhou, Haotian Fu, Yulong Huang, Bojun Cheng:
A Simple and Effective Point-Based Network for Event Camera 6-DOFs Pose Relocalization. 18112-18121
- Shangjie Xue, Jesse Dill, Pranay Mathur, Frank Dellaert, Panagiotis Tsotras, Danfei Xu:
Neural Visibility Field for Uncertainty-Driven Active Mapping. 18122-18132
- Shagun Uppal, Ananya Agarwal, Haoyu Xiong, Kenneth Shaw, Deepak Pathak:
SPIN: Simultaneous Perception, Interaction and Navigation. 18133-18142
- Xuesong Nie, Haoyuan Jin, Yunfeng Yan, Xi Chen, Zhihang Zhu, Donglian Qi:
PredToken: Predicting Unknown Tokens and Beyond with Coarse-to-Fine Iterative Decoding. 18143-18152
- Jacob Chalk, Jaesung Huh, Evangelos Kazakos, Andrew Zisserman, Dima Damen:
TIM: A Time Interval Machine for Audio-Visual Action Recognition. 18153-18163
- Tengda Han, Max Bain, Arsha Nagrani, Gü̈l Varol, Weidi Xie, Andrew Zisserman:
AutoAD III: The Prequel - Back to the Pixels. 18164-18174
- Zijia Lu, Ehsan Elhamifar:
FACT: Frame-Action Cross-Attention Temporal Modeling for Efficient Action Segmentation. 18175-18185
- Yuhan Shen, Ehsan Elhamifar:
Progress-Aware Online Action Segmentation for Egocentric Procedural Task Videos. 18186-18197
- Md Mohaiminul Islam, Ngan Ho, Xitong Yang, Tushar Nagarajan, Lorenzo Torresani, Gedas Bertasius:
Video ReCap: Recursive Captioning of Hour-Long Videos. 18198-18208
- Junke Wang, Dongdong Chen, Chong Luo, Bo He, Lu Yuan, Zuxuan Wu, Yu-Gang Jiang:
OmniViD: A Generative Framework for Universal Video Understanding. 18209-18220
- Enxin Song, Wenhao Chai, Guanhong Wang, Yucheng Zhang, Haoyang Zhou, Feiyang Wu, Haozhe Chi, Xun Guo, Tian Ye, Yanting Zhang, Yan Lu, Jenq-Neng Hwang, Gaoang Wang:
MovieChat: From Dense Token to Sparse Memory for Long Video Understanding. 18221-18232
- Chihiro Nakatani, Hiroaki Kawashima, Norimichi Ukita:
Learning Group Activity Features Through Person Attribute Prediction. 18233-18242
- Xingyi Zhou, Anurag Arnab, Shyamal Buch, Shen Yan, Austin Myers, Xuehan Xiong, Arsha Nagrani, Cordelia Schmid:
Streaming Dense Video Captioning. 18243-18252



- Angchi Xu, Wei-Shi Zheng:
Efficient and Effective Weakly-Supervised Action Segmentation via Action-Transition-Aware Boundary Alignment. 18253-18262
- Runhao Zeng, Xiaoyong Chen, Jiaming Liang, Huisi Wu, Guangzhong Cao, Yong Guo:
Benchmarking the Robustness of Temporal Action Detection Models Against Temporal Corruptions. 18263-18274
- Simone Alberto Peirone, Francesca Pistilli, Antonio Alliegro, Giuseppe Averta:
A Backpack Full of Skills: Egocentric Video Understanding with Diverse Task Perspectives. 18275-18285
- Razvan-George Pasca, Alexey Gavryushin, Muhammad Hamza, Yen-Ling Kuo, Kaichun Mo, Luc Van Gool, Otmar Hilliges, Xi Wang:
Summarize the Past to Predict the Future: Natural Language Descriptions of Context Boost Multimodal Object Interaction Anticipation. 18286-18296
- Peng Wu, Xuerong Zhou, Guansong Pang, Yujia Sun, Jing Liu, Peng Wang, Yanning Zhang:
Open-Vocabulary Video Anomaly Detection. 18297-18307
- Jin Yang, Ping Wei, Huan Li, Ziyang Ren:
Task-Driven Exploration: Decoupling and Inter-Task Feedback for Joint Moment Retrieval and Highlight Detection. 18308-18318
- Junxi Chen, Liang Li, Li Su, Zheng-Jun Zha, Qingming Huang:
Prompt-Enhanced Multiple Instance Learning for Weakly Supervised Video Anomaly Detection. 18319-18329
- Xin Gu, Heng Fan, Yan Huang, Tiejian Luo, Libo Zhang:
Context-Guided Spatio-Temporal Video Grounding. 18330-18339
- Dominick Reilly, Srijan Das:
Just Add π! Pose Induced Video Transformers for Understanding Activities of Daily Living. 18340-18350
- Lin Geng Foo, Tianjiao Li, Hossein Rahmani, Jun Liu:
Action Detection via an Image Diffusion Process. 18351-18361
- Jia Gong, Lin Geng Foo, Yixuan He, Hossein Rahmani, Jun Liu:
LLMs are Good Sign Language Translators. 18362-18372
- Alexey A. Gritsenko, Xuehan Xiong, Josip Djolonga, Mostafa Dehghani, Chen Sun, Mario Lucic, Cordelia Schmid, Anurag Arnab:
End-to-End Spatio-Temporal Action Localisation with Video Transformers. 18373-18383
- Trong-Thuan Nguyen, Pha A. Nguyen, Khoa Luu:
HIG: Hierarchical Interlacement Graph Approach to Scene Graph Generation in Video Understanding. 18384-18394
- Haoxuan Qu, Yujun Cai, Jun Liu:
LLMs are Good Action Recognizers. 18395-18406
- Joya Chen, Zhaoyang Lv, Shiwei Wu, Kevin Qinghong Lin, Chenan Song, Difei Gao, Jia-Wei Liu, Ziteng Gao, Dongxing Mao, Mike Zheng Shou:
VideoLLM-online: Online Video Large Language Model for Streaming Video. 18407-18418
- Brian Chen, Nina Shvetsova, Andrew Rouditchenko, Daniel Kondermann, Samuel Thomas, Shih-Fu Chang, Rogério Feris, James R. Glass, Hilde Kuehne:
What, When, and Where? Self-Supervised Spatio-Temporal Grounding in Untrimmed Multi-Action Videos from Narrated Instructions. 18419-18429
- Shiyi Zhang, Sule Bai, Guangyi Chen, Lei Chen, Jiwen Lu, Junle Wang, Yansong Tang:
Narrative Action Evaluation with Prompt-Guided Multimodal Interaction. 18430-18439
- Ziying Xia, Jian Cheng, Siyu Liu, Yongxiang Hu, Shiguang Wang, Yijie Zhang, Liwan Dang:
Realigning Confidence with Temporal Saliency Information for Point-Level Weakly-Supervised Temporal Action Localization. 18440-18450
- Chi-Hsi Kung, Shu-Wei Lu, Yi-Hsuan Tsai, Yi-Ting Chen:
Action-Slot: Visual Action-Centric Representations for Multi-Label Atomic Activity Recognition in Traffic Scenes. 18451-18461
- Xizi Wang, Feng Cheng, Gedas Bertasius:
LoCoNet: Long-Short Context Network for Active Speaker Detection. 18462-18472
- Jiawei Tan, Hongxing Wang, Jiaxin Li, Zhilong Ou, Zhangbin Qian:
Neighbor Relations Matter in Video Scene Detection. 18473-18482
- Alessandro Flaborea, Guido Maria D'Amely di Melendugno, Leonardo Plini, Luca Scofano, Edoardo De Matteis, Antonino Furnari, Giovanni Maria Farinella, Fabio Galasso:
PREGO: Online Mistake Detection in PROcedural EGOCentric Videos. 18483-18492
- Zihui Xue, Kumar Ashutosh, Kristen Grauman:
Learning Object State Changes in Videos: An Open-World Perspective. 18493-18503
- Wei Zhang, Chaoqun Wan, Tongliang Liu, Xinmei Tian, Xu Shen, Jieping Ye:
Enhanced Motion-Text Alignment for Image-to-Video Transfer Learning. 18504-18515
- Zhiyu Zhao, Bingkun Huang, Sen Xing, Gangshan Wu, Yu Qiao, Limin Wang:
Asymmetric Masked Distillation for Pre-Training Small Foundation Models. 18516-18526



- Luca Zanella, Willi Menapace, Massimiliano Mancini, Yiming Wang, Elisa Ricci:
Harnessing Large Language Models for Training-Free Video Anomaly Detection. 18527-18536
- Tao Wu, Runyu He, Gangshan Wu, Limin Wang:
SportsHHI: A Dataset for Human-Human Interaction Detection in Sports Videos. 18537-18546
- Kumara Kahatapitiya, Anurag Arnab, Arsha Nagrani, Michael S. Ryoo:
VicTR: Video-conditioned Text Representations for Activity Recognition. 18547-18558
- Yuhan Zhu, Guozhen Zhang, Jing Tan, Gangshan Wu, Limin Wang:
Dual DETRs for Multi-Label Temporal Action Detection. 18559-18569
- Min Yang, Huan Gao, Ping Guo, Limin Wang:
Adapting Short-Term Transformers for Action Detection in Untrimmed Videos. 18570-18579
- Himangi Mittal, Nakul Agarwal, Shao-Yuan Lo, Kwonjoon Lee:
Can't make an Omelette without Breaking some Eggs: Plausible Action Anticipation using Large Video-Language Models. 18580-18590
- Shuming Liu, Chen-Lin Zhang, Chen Zhao, Bernard Ghanem:
End-to-End Temporal Action Detection with 1B Parameters Across 1000 Frames. 18591-18601
- Junbao Zhou, Ziqi Pang, Yu-Xiong Wang:
RMem: Restricted Memory Banks Improve Video Object Segmentation. 18602-18611
- Friedhelm Hamann, Suman Ghosh, Ignacio Juarez Martinez, Tom Hart, Alex Kacelnik, Guillermo Gallego:
Low-power, Continuous Remote Behavioral Localization with Event Cameras. 18612-18621
- Ivan Rodin, Antonino Furnari, Kyle Min, Subarna Tripathi, Giovanni Maria Farinella:
Action Scene Graphs for Long-Form Understanding of Egocentric Videos. 18622-18632
- Jiazhou Zhou, Xu Zheng, Yuanhuiyi Lyu, Lin Wang:
ExACT: Language-Guided Conceptual Reasoning and Uncertainty Estimation for Event-Based Action Recognition and More. 18633-18643
- Hongji Guo, Nakul Agarwal, Shao-Yuan Lo, Kwonjoon Lee, Qiang Ji:
Uncertainty-aware Action Decoupling Transformer for Action Anticipation. 18644-18654
- Shih-Po Lee, Zijia Lu, Zekun Zhang, Minh Hoai, Ehsan Elhamifar:
Error Detection in Egocentric Procedural Task Videos. 18655-18666
- Gerard Donahue, Ehsan Elhamifar:
Learning to Predict Activity Progress by Self-Supervised Video Alignment. 18667-18677
- Mohamed Abdelfattah, Mariam Hassan, Alexandre Alahi:
MaskCLR: Attention-Guided Contrastive Learning for Robust Action Representation Learning. 18678-18687
- Yifei Chen, Dapeng Chen, Ruijin Liu, Sai Zhou, Wenyuan Xue, Wei Peng:
Align Before Adapt: Leveraging Entity-to-Region Alignments for Generalizable Video Action Recognition. 18688-18698
- Hao Wu, Huabin Liu, Yu Qiao, Xiao Sun:
DIBS: Enhancing Dense Video Captioning with Unlabeled Videos via Pseudo Boundary Enrichment and Online Refinement. 18699-18708
- Yicheng Xiao, Zhuoyan Luo, Yong Liu, Yue Ma, Hengwei Bian, Yatai Ji, Yujiu Yang, Xiu Li:
Bridging the Gap: A Unified Video Comprehension Framework for Moment Retrieval and Highlight Detection. 18709-18719
- Benedetta Liberatori, Alessandro Conti, Paolo Rota, Yiming Wang, Elisa Ricci:
Test-Time Zero-Shot Temporal Action Localization. 18720-18729
- Filip Illic, He Zhao, Thomas Pock, Richard P. Wildes:
Selective, Interpretable and Motion Consistent Privacy Attribute Obfuscation for Action Recognition. 18730-18739
- Tushar Nagarajan, Lorenzo Torresani:
Step Differences in Instructional Video. 18740-18750
- Hoyeoung Yun, Jinwoo Ahn, Minseo Kim, Eun-Sol Kim:
Compositional Video Understanding with Spatiotemporal Structure-based Transformers. 18751-18760
- Anqi Zhu, Qiuhong Ke, Mingming Gong, James Bailey:
Part-Aware Unified Representation of Language and Skeleton for Zero-Shot Action Recognition. 18761-18770
- Joonmyung Choi, Sanghyeok Lee, Jaewon Chu, Minhyuk Choi, Hyunwoo J. Kim:
vid-TLDR: Training Free Token merging for Light-Weight Video Transformer. 18771-18781
- Shunli Wang, Shuaibing Wang, Dingkang Yang, Mingcheng Li, Haopeng Kuang, Xiao Zhao, Liuzhen Su, Peng Zhai, Lihua Zhang:
CPR-Coach: Recognizing Composite Error Actions Based on Single-Class Training. 18782-18792
- Hang Du, Sicheng Zhang, Binzhu Xie, Guoshun Nan, Jiayang Zhang, Junrui Xu, Hangyu Liu, Sicong Leng, Jiangming Liu, Hehe Fan, Dajiu Huang, Jing Feng, Linli Chen, Can Zhang, Xuhuan Li, Hao Zhang, Jianhang Chen, Qimei Cui, Xiaofeng Tao:
Uncovering what, why and How: A Comprehensive Benchmark for Causation Understanding of Video Anomaly. 18793-18803



- Kumar Ashutosh, Zihui Xue, Tushar Nagarajan, Kristen Grauman:
Detours for Navigating Instructional Videos. 18804-18815
- Kumaranage Ravindu Yasaras Nagasinghe, Honglu Zhou, Malitha Gunawardhana, Martin Renqiang Min, Daniel Harari, Muhammad Haris Khan:
Why Not Use Your Textbook? Knowledge-Enhanced Procedure Planning of Instructional Videos. 18816-18826
- Ioanna Ntinou, Enrique Sanchez, Georgios Tzimiropoulos:
Multiscale Vision Transformers Meet Bipartite Matching for Efficient Single-Stage Action Localization. 18827-18836
- Ho-Joong Kim, Jung-Ho Hong, Heejo Kong, Seong-Whan Lee:
TE-TAD: Towards Full End-to-End Temporal Action Detection via Time-Aligned Coordinate Expression. 18837-18846
- Jaewon Son, Jaehun Park, Kwangsu Kim:
CSTA: CNN-based Spatiotemporal Attention for Video Summarization. 18847-18856
- Haosong Zhang, Mei Chee Leong, Liyuan Li, Weisi Lin:
PeVL: Pose-Enhanced Vision-Language Model for Fine-Grained Human Action Recognition. 18857-18867
- Jakub Micorek, Horst Possegger, Dominik Narnhofer, Horst Bischof, Mateusz Kozinski:
MULDE: Multiscale Log-Density Estimation via Denoising Score Matching for Video Anomaly Detection. 18868-18877
- Ning Wang, Guangming Zhu, HS Li, Liang Zhang, Syed Afaq Ali Shah, Mohammed Bennamoun:
Language Model Guided Interpretable Video Action Reasoning. 18878-18887
- Tom Tongjia Chen, Hongshan Yu, Zhengeng Yang, Zechuan Li, Wei Sun, Chen Chen:
OST: Refining Text Knowledge with Optimal Spatio-Temporal Descriptor for General Video Recognition. 18888-18898
- Zhiwei Yang, Jing Liu, Peng Wu:
Text Prompt with Normality Guidance for Weakly Supervised Video Anomaly Detection. 18899-18908
- Syed Talal Wasim, Muzammal Naseer, Salman H. Khan, Ming-Hsuan Yang, Fahad Shahbaz Khan:
VideoGrounding-DINO: Towards Open-Vocabulary Spatio-Temporal Video Grounding. 18909-18918
- Arun V. Reddy, William Paul, Corban Rivera, Ketul Shah, Celso M. de Melo, Rama Chellappa:
Unsupervised Video Domain Adaptation with Masked Pre-Training and Collaborative Self-Training. 18919-18929
- Fangzhou Mu, Sicheng Mo, Yin Li:
SnAG: Scalable and Accurate Video Grounding. 18930-18940
- Manjin Kim, Paul Hongsuck Seo, Cordelia Schmid, Minsu Cho:
Learning Correlation Structures for Vision Transformers. 18941-18951
- Kranthi Kumar Rachavarapu, Kalyan Ramakrishnan, A. N. Rajagopalan:
Weakly-Supervised Audio-Visual Video Parsing with Prototype-Based Pseudo-Labeling. 18952-18962
- Siyuan Li, Lei Ke, Martin Danelljan, Luigi Piccinelli, Mattia Segù, Luc Van Gool, Fisher Yu:
Matching Anything by Segmenting Anything. 18963-18973
- Siqi Li, Zhikuan Zhou, Zhou Xue, Yipeng Li, Shaoyi Du, Yue Gao:
3D Feature Tracking via Event Camera. 18974-18983
- Fei Wang, Dan Guo, Kun Li, Zhun Zhong, Meng Wang:
Frequency Decoupling for Motion Magnification Via Multi-Level Isomorphic Architecture. 18984-18994
- Zheng Qin, Le Wang, Sanping Zhou, Panpan Fu, Gang Hua, Wei Tang:
Towards Generalizable Multi-Object Tracking. 18995-19004
- Conghao Wong, Beihao Xia, Ziqian Zou, Yulong Wang, Xinge You:
SocialCircle: Learning the Angle-based Social Interaction Representation for Pedestrian Trajectory Prediction. 19005-19015
- Zijia Lu, Bing Shuai, Yanbei Chen, Zhenlin Xu, Davide Modolo:
Self-Supervised Multi-Object Tracking with Path Consistency. 19016-19026
- Shuai Yuan, Lei Luo, Zhuo Hui, Can Pu, Xiaoyu Xiang, Rakesh Ranjan, Denis Demandolx:
UnSAMFlow: Unsupervised Optical Flow Guided by Segment Anything Model. 19027-19037
- Yuqing Huang, Xin Li, Zikun Zhou, Yaowei Wang, Zhenyu He, Ming-Hsuan Yang:
RTracker: Recoverable Tracking via PN Tree Structured Memory. 19038-19047
- Yifan Bai, Zeyang Zhao, Yihong Gong, Xing Wei:
ARTrackV2: Prompting Autoregressive Tracker Where to Look and How to Describe. 19048-19057
- Wenjun Hui, Zhenfeng Zhu, Shuai Zheng, Yao Zhao:
Endow SAM with Keen Eyes: Temporal-Spatial Prompt Learning for Video Camouflaged Object Detection. 19058-19067
- Qiaole Dong, Yanwei Fu:
MemFlow: Optical Flow Estimation and Prediction with Memory. 19068-19078
- Lingyi Hong, Shilin Yan, Renrui Zhang, Wanyun Li, Xinyu Zhou, Pinxue Guo, Kaixun Jiang, Yiting Chen, Jinglun Li, Zhaoyu Chen, Wenqiang Zhang:
OneTracker: Unifying Visual Object Tracking with Foundation Models and Efficient Tuning. 19079-19091



- Yaroslava Lochman, Carl Olsson, Christopher Zach:
Learned Trajectory Embedding for Subspace Clustering. 19092-19102
- Qi Zhao, M. Salman Asif, Zhan Ma:
PNeRV: Enhancing Spatial Consistency via Pyramidal Neural Representation for Videos. 19103-19112
- Fei Xie, Zhongdao Wang, Chao Ma:
DiffusionTrack: Point Set Diffusion Model for Visual Object Tracking. 19113-19124
- Chunxu Liu, Guozhen Zhang, Rui Zhao, Limin Wang:
Sparse Global Matching for Video Frame Interpolation with Large Motion. 19125-19134
- Yunhao Du, Cheng Lei, Zhicheng Zhao, Fei Su:
iKUN: Speak to Trackers Without Retraining. 19135-19144
- Guangze Zheng, Shijie Lin, Haobo Zuo, Changhong Fu, Jia Pan:
NetTrack: Tracking Highly Dynamic Objects with a Net. 19145-19155
- Zongwei Wu, Jilai Zheng, Xiangxuan Ren, Florin-Alexandru Vasluiianu, Chao Ma, Danda Pani Paudel, Luc Van Gool, Radu Timofte:
Single-Model and Any-Modality for Video Object Tracking. 19156-19166
- Ao Luo, Xin Li, Fan Yang, Jiangyu Liu, Haoqiang Fan, Shuaicheng Liu:
FlowDiffuser: Advancing Optical Flow Estimation with Diffusion Models. 19167-19176
- Zonghui Guo, Xinyu Han, Jie Zhang, Shiguang Shan, Haiyong Zheng:
Video Harmonization with Triplet Spatio-Temporal Variation Patterns. 19177-19186
- Guillaume Le Moing, Jean Ponce, Cordelia Schmid:
Dense Optical Tracking: Connecting the Dots. 19187-19197
- Xinglong Luo, Ao Luo, Zhengning Wang, Chunyu Lin, Bing Zeng, Shuaicheng Liu:
Efficient Meshflow and Optical Flow Estimation from Event Cameras. 19198-19207
- Yanyan Shao, Shuteng He, Qi Ye, Yuchao Feng, Wenhan Luo, Jiming Chen:
Context-Aware Integration of Language and Visual References for Natural Language Tracking. 19208-19217
- Weihuang Liu, Xi Shen, Haolun Li, Xiuli Bi, Bo Liu, Chi-Man Pun, Xiaodong Cun:
Depth-Aware Test-Time Training for Zero-Shot Video Object Segmentation. 19218-19227
- Xinyan Liu, Guorong Li, Yuankai Qi, Ziheng Yan, Zhenjun Han, Anton van den Hengel, Ming-Hsuan Yang, Qingming Huang:
Weakly Supervised Video Individual Counting. 19228-19237
- Suhwan Cho, Minhyeok Lee, Seunghoon Lee, Dogyoon Lee, Heeseung Choi, Ig-Jae Kim, Sangyoun Lee:
Dual Prototype Attention for Unsupervised Video Object Segmentation. 19238-19247
- Xiao Wang, Shiao Wang, Chuanming Tang, Lin Zhu, Bo Jiang, Yonghong Tian, Jin Tang:
Event Stream-Based Visual Object Tracking: A High-Resolution Benchmark Dataset and A Novel Baseline. 19248-19257
- Wenrui Cai, Qingjie Liu, Yunhong Wang:
HIPTrack: Visual Tracking with Historical Prompts. 19258-19267
- Seokju Cho, Jiahui Huang, Seungryong Kim, Joon-Young Lee:
FlowTrack: Revisiting Optical Flow for Long-Range Dense Tracking. 19268-19277
- Yue Gao, Jiahao Li, Lei Chu, Yan Lu:
Implicit Motion Function. 19278-19289
- Cheng Huang, Shoudong Han, Mengyu He, Wenbo Zheng, Yuhao Wei:
DeconfuseTrack: Dealing with Confusion for Multi-Object Tracking. 19290-19299
- Jinxia Xie, Bineng Zhong, Zhiyi Mo, Shengping Zhang, Liangtao Shi, Shuxiang Song, Rongrong Ji:
Autoregressive Queries for Adaptive Tracking with Spatio-Temporal Transformers. 19300-19309
- Zhicheng Zhang, Junyao Hu, Wentao Cheng, Danda Paudel, Jufeng Yang:
ExtDM: Distribution Extrapolation Diffusion Model for Video Prediction. 19310-19320
- Weiyi Lv, Yuhang Huang, Ning Zhang, Ruei-Sung Lin, Mei Han, Dan Zeng:
DiffMOT: A Real-time Diffusion-based Multiple Object Tracker with Non-linear Prediction. 19321-19330
- Haozhe Lin, Chunyu Wei, Li He, Yuchen Guo, Yunqi Zhao, Shanglong Li, Lu Fang:
GigaTraj: Predicting Long-term Trajectories of Hundreds of Pedestrians in Gigapixel Complex Scenes. 19331-19340
- Sijia Chen, En Yu, Jinyang Li, Wenbing Tao:
Delving into the Trajectory Long-tail Distribution for Multi-object Tracking. 19341-19351
- Jisoo Jeong, Hong Cai, Risheek Garrepalli, Jamie Menjay Lin, Munawar Hayat, Fatih Porikli:
OCAI: Improving Optical Flow Estimation by Occlusion and Consistency Aware Interpolation. 19352-19362
- Yuanbang Liang, Bhavesh Garg, Paul L. Rosin, Yipeng Qin:
Deep Generative Model based Rate-Distortion for Image Downscaling Assessment. 19363-19372
- Hao Chen, Yuqi Hou, Chenyuan Qu, Irene Testini, Xiaohan Hong, Jianbo Jiao:
360+x: A Panoptic Multi-modal Scene Understanding Dataset. 19373-19382
- Kristen Grauman, Andrew Westbury, Lorenzo Torresani, Kris Kitani, Jitendra Malik, Triantafyllos Afouras, Kumar Ashutosh, Vijay Baiyya, Siddhant Bansal, Bikram Boote, Eugene Byrne, Zachary Chavis, Joya Chen, Feng Cheng, Fu-Jen Chu, Sean



Crane, Avijit Dasgupta, Jing Dong, María Escobar, Cristhian Forigua, Abrham Gebreselasie, Sanjay Haresh, Jing Huang, Md Mohaiminul Islam, Suyog Dutt Jain, Rawal Khirodkar, Devansh Kukreja, Kevin J. Liang, Jia-Wei Liu, Sagnik Majumder, Yongsen Mao, Miguel Martin, Effrosyni Mavroudi, Tushar Nagarajan, Francesco Ragusa, Santhosh Kumar Ramakrishnan, Luigi Seminara, Arjun Somayazulu, Yale Song, Shan Su, Zihui Xue, Edward Zhang, Jinxu Zhang, Angela Castillo, Changan Chen, Xinzhu Fu, Ryosuke Furuta, Cristina González, Prince Gupta, Jiabo Hu, Yifei Huang, Yiming Huang, Weslie Khoo, Anush Kumar, Robert Kuo, Sach Lakhavani, Miao Liu, Mi Luo, Zhengyi Luo, Brighid Meredith, Austin Miller, Oluwatumininu Oguntola, Xiaqing Pan, Penny Peng, Shraman Pramanick, Merey Ramazanova, Fiona Ryan, Wei Shan, Kiran Somasundaram, Chenan Song, Audrey Southerland, Masatoshi Tateno, Huiyu Wang, Yuchen Wang, Takuma Yagi, Mingfei Yan, Xitong Yang, Zecheng Yu, Shengxin Cindy Zha, Chen Zhao, Ziwei Zhao, Zhifan Zhu, Jeff Zhuo, Pablo Arbeláez, Gedas Bertasius, Dima Damen, Jakob Engel, Giovanni Maria Farinella, Antonino Furnari, Bernard Ghanem, Judy Hoffman, C. V. Jawahar, Richard A. Newcombe, Hyun Soo Park, James M. Rehg, Yoichi Sato, Manolis Savva, Jianbo Shi, Mike Zheng Shout, Michael Wray:

Ego-Exo4D: Understanding Skilled Human Activity from First- and Third-Person Perspectives. 19383-19400

- Youwei Liang, Junfeng He, Gang Li, Peizhao Li, Arseniy Klimovskiy, Nicholas Carolan, Jiao Sun, Jordi Pont-Tuset, Sarah Young, Feng Yang, Junjie Ke, Krishnamurthy Dj Dvijotham, Katherine M. Collins, Yiwen Luo, Yang Li, Kai J. Kohlhoff, Deepak Ramachandran, Vidhya Navalpakkam:

Rich Human Feedback for Text-to-Image Generation. 19401-19411

- Samuel Stevens, Jiaman Wu, Matthew J. Thompson, Elizabeth G. Campolongo, Chan Hee Song, David Edward Carolyn, Li Dong, Wasila M. Dahdul, Charles V. Stewart, Tanya Y. Berger-Wolf, Wei-Lun Chao, Yu Su:

BioCLIP: A Vision Foundation Model for the Tree of Life. 19412-19424

- Zelin Zhao, Fenglei Fan, Wenlong Liao, Junchi Yan:

Grounding and Enhancing Grid-based Models for Neural Fields. 19425-19435

- Jiahao Chen, Yipeng Qin, Lingjie Liu, Jiangbo Lu, Guanbin Li:

NeRF-HuGS: Improved Neural Radiance Fields in Non-static Scenes Using Heuristics-Guided Segmentation. 19436-19446

- Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, Andreas Geiger:

Mip-Splatting: Alias-Free 3D Gaussian Splatting. 19447-19456

- David Charatan, Sizhe Lester Li, Andrea Tagliasacchi, Vincent Sitzmann:

PixelSplat: 3D Gaussian Splats from Image Pairs for Scalable Generalizable 3D Reconstruction. 19457-19467

- Khang Truong Giang, Soohwan Song, Sungho Jo:

Learning to Produce Semi-Dense Correspondences for Visual Localization. 19468-19478

- Shiyu Tian, Hongxin Wei, Yiqun Wang, Lei Feng:

CroSel: Cross Selection of Confident Pseudo Labels for Partial-Label Learning. 19479-19488

- Siyao Lin, Pumeng Lyu, Dongrui Liu, Tao Tang, Xiaodan Liang, Andy Song, Xiaojun Chang:

MLP Can Be a Good Transformer Learner. 19489-19498

- Hyeokjun Kweon, Kuk-Jin Yoon:

From SAM to CAMs: Exploring Segment Anything Model for Weakly Supervised Semantic Segmentation. 19499-19509

- Qihao Zhao, Yalun Dai, Hao Li, Wei Hu, Fan Zhang, Jun Liu:

LTGC: Long-Tail Recognition via Leveraging LLMs-Driven Generated Content. 19510-19520

- Octave Mariotti, Oisin Mac Aodha, Hakan Bilen:

Improving Semantic Correspondence with Viewpoint-Guided Spherical Maps. 19521-19530

- Xuying Zhang, Bowen Yin, Yuming Chen, Zheng Lin, Yunheng Li, Qibin Hou, Ming-Ming Cheng:

TeMO: Towards Text-Driven 3D Stylization for Multi-Object Meshes. 19531-19540

- Ethan Elms, Yasir Latif, Tae Ha Park, Tat-Jun Chin:

Event-based Structure-from-Orbit. 19541-19550

- Xiaoyang Wu, Zhuotao Tian, Xin Wen, Bohao Peng, Xihui Liu, Kaicheng Yu, Hengshuang Zhao:

Towards Large-Scale 3D Representation Learning with Multi-Dataset Point Prompt Training. 19551-19562

- Shanlin Sun, Bingbing Zhuang, Ziyu Jiang, Buyu Liu, Xiaohui Xie, Manmohan Chandraker:

LidaRF: Delving into Lidar for Neural Radiance Field on Street Scenes. 19563-19572

- Di Liu, Bingbing Zhuang, Dimitris N. Metaxas, Manmohan Chandraker:

Instantaneous Perception of Moving Objects in 3D. 19573-19583

- Delin Qu, Chi Yan, Dong Wang, Jie Yin, Qizhi Chen, Dan Xu, Yiting Zhang, Bin Zhao, Xuelong Li:

Implicit Event-RGBD Neural SLAM. 19584-19594

- Chi Yan, Delin Qu, Dan Xu, Bin Zhao, Zhigang Wang, Dong Wang, Xuelong Li:

GS-SLAM: Dense Visual SLAM with 3D Gaussian Splatting. 19595-19604

- Zhiyuan Yu, Zheng Qin, Lintao Zheng, Kai Xu:

Learning Instance-Aware Correspondences for Robust Multi-Instance Point Cloud Registration in Cluttered Scenes. 19605-19614

- Yawar Siddiqui, Antonio Alliegro, Alexey Artemov, Tatiana Tommasi, Daniele Sirigatti, Vladislav Rosov, Angela Dai, Matthias Nießner:

MeshGPT: Generating Triangle Meshes with Decoder-Only Transformers. 19615-19625



- Lahav Lipson, Jia Deng:
Multi-Session SLAM with Differentiable Wide-Baseline Pose Optimization. 19626-19635
- Andreas Engelhardt, Amit Raj, Mark Boss, Yunzhi Zhang, Abhishek Kar, Yuanzhen Li, Deqing Sun, Ricardo Martin-Brualla, Jonathan T. Barron, Hendrik P. A. Lensch, Varun Jampani:
SHINOBI: Shape and Illumination using Neural Object Decomposition via BRDF Optimization In-the-wild. 19636-19646
- Haithem Turki, Vasu Agrawal, Samuel Rota Bulò, Lorenzo Porzi, Peter Kotschieder, Deva Ramanan, Michael Zollhöfer, Christian Richardt:
HybridNeRF: Efficient Neural Rendering via Adaptive Volumetric Surfaces. 19647-19656
- Tianchen Deng, Guole Shen, Tong Qin, Jianyu Wang, Wentao Zhao, Jingchuan Wang, Danwei Wang, Weidong Chen:
PLGSLAM: Progressive Neural Scene Represenation with Local to Global Bundle Adjustment. 19657-19666
- Xinhang Liu, Yu-Wing Tai, Chi-Keung Tang, Pedro Miraldo, Suhas Lohit, Moitreyaa Chatterjee:
Gear-NeRF: Free-Viewpoint Rendering and Tracking with Motion-Aware Spatio-Temporal Sampling. 19667-19679
- Shunyuan Zheng, Boyao Zhou, Ruizhi Shao, Boning Liu, Shengping Zhang, Liqiang Nie, Yebin Liu:
GPS-Gaussian: Generalizable Pixel-Wise 3D Gaussian Splatting for Real-Time Human Novel View Synthesis. 19680-19690
- Zhiying Leng, Tolga Birdal, Xiaohui Liang , Federico Tombari:
HyperSDFusion: Bridging Hierarchical Structures in Language and Geometry for Enhanced 3D Text2Shape Generation. 19691-19700
- Xianqi Wang, Gangwei Xu, Hao Jia, Xin Yang:
Selective-Stereo: Adaptive Frequency Information Selection for Stereo Matching. 19701-19710
- Zhe Li, Zerong Zheng, Lizhen Wang, Yebin Liu:
Animatable Gaussians: Learning Pose-Dependent Gaussian Maps for High-Fidelity Human Avatar Modeling. 19711-19722
- Thomas Tanay, Matteo Maggioni:
Global Latent Neural Rendering. 19723-19733
- Yuheng Jiang, Zhehao Shen, Penghao Wang, Zhuo Su, Yu Hong, Yingliang Zhang, Jingyi Yu, Lan Xu:
HiFi4G: High-Fidelity Human Performance Rendering via Compact Gaussian Splatting. 19734-19745
- Kunhong Li, Longguang Wang, Ye Zhang, Kaiwen Xue, Shunbo Zhou, Yulan Guo:
LoS: Local Structure-Guided Stereo Matching. 19746-19756
- Tai Wang, Xiaohan Mao, Chenming Zhu, Runsen Xu, Ruiyuan Lyu, Peisen Li, Xiao Chen, Wenwei Zhang, Kai Chen, Tianfan Xue, Xihui Liu, Cewu Lu, Dahua Lin, Jiangmiao Pang:
EmbodiedScan: A Holistic Multi-Modal 3D Perception Suite Towards Embodied AI. 19757-19767
- Jinyoung Jun, Jae-Han Lee, Chang-Su Kim:
Masked Spatial Propagation Network for Sparsity-Adaptive Depth Refinement. 19768-19778
- Yuanmin Huang, Mi Zhang, Daizong Ding, Erling Jiang, Zhaoxiang Wang, Min Yang:
CausalPC: Improving the Robustness of Point Cloud Classification by Causal Effect Identification. 19779-19789
- Johan Edstedt, Qiyu Sun, Georg Bökman, Mårten Wadenbäck, Michael Felsberg:
RoMa: Robust Dense Feature Matching. 19790-19800
- Zhangyang Xiong, Chenghong Li, Kenkun Liu, Hongjie Liao, Jianqiao Hu, Junyi Zhu, Shuliang Ning, Lingteng Qiu, Chongjie Wang, Shijie Wang, Shuguang Cui, Xiaoguang Han:
MVHumanNet: A Large-Scale Dataset of Multi-View Daily Dressing Human Captures. 19801-19811
- Abdullah Hamdi, Luke Melas-Kyriazi, Jinjie Mai, Guocheng Qian, Ruoshi Liu, Carl Vondrick, Bernard Ghanem, Andrea Vedaldi:
GES: Generalized Exponential Splatting for Efficient Radiance Field Rendering. 19812-19822
- Jihan Yang, Runyu Ding, Weipeng Deng, Zhe Wang, Xiaojuan Qi:
RegionPLC: Regional Point-Language Contrastive Learning for Open-World 3D Scene Understanding. 19823-19832
- Zinuo You, Andreas Geiger, Anpei Chen:
NeLF-Pro: Neural Light Field Probes for Multi-Scale Novel View Synthesis. 19833-19843
- Weirong Chen, Le Chen, Rui Wang, Marc Pollefeys:
LEAP-VO: Long-term Effective Any Point Tracking for Visual Odometry. 19844-19853
- Chris Rockwell, Nilesh Kulkarni, Linyi Jin, Jeong Joon Park, Justin Johnson, David F. Fouhey:
FAR: Flexible, Accurate and Robust 6DoF Relative Camera Pose Estimation. 19854-19864
- Hanwen Jiang, Arjun Karpur, Bingyi Cao, Qixing Huang, André Araújo:
OmniGlue: Generalizable Feature Matching with Foundation Model Guidance. 19865-19875
- Jiahui Lei, Yufu Wang, Georgios Pavlakos, Lingjie Liu, Kostas Daniilidis:
GART: Gaussian Articulated Template Models. 19876-19887
- Christian Diller, Angela Dai:
CG-HOI: Contact-Guided 3D Human-Object Interaction Generation. 19888-19901
- Christian Diller, Thomas A. Funkhouser, Angela Dai:
FutureHuman3D: Forecasting Complex Long-Term 3D Human Behavior from Video Observations. 19902-19914



- Ying-Tian Liu, Yuan-Chen Guo, Guan Luo, Heyi Sun, Wei Yin, Song-Hai Zhang:
PI3D: Efficient Text-to-3D Generation with Pseudo-Image Diffusion. 19915-19924
- Haoming Chen, Zhizhong Zhang, Yanyun Qu, Ruixin Zhang, Xin Tan, Yuan Xie:
Building a Strong Pre-Training Baseline for Universal 3D Large-Scale Perception. 19925-19935
- Qihang Ma, Xin Tan, Yanyun Qu, Lizhuang Ma, Zhizhong Zhang, Yuan Xie:
COTR: Compact Occupancy TRansformer for Vision-Based 3D Occupancy Prediction. 19936-19945
- Yuanhui Huang, Wenzhao Zheng, Borui Zhang, Jie Zhou, Jiwen Lu:
SelfOcc: Self-Supervised Vision-Based 3D Occupancy Prediction. 19946-19956
- Dávid Rozenberszki, Or Litany, Angela Dai:
UnScene3D: Unsupervised 3D Instance Segmentation for Indoor Scenes. 19957-19967
- Nan Xue, Bin Tan, Yuxi Xiao, Liang Dong, Gui-Song Xia, Tianfu Wu, Yujun Shen:
NEAT: Distilling 3D Wireframes from Neural Attraction Fields. 19968-19977
- Yizhak Ben-Shabat, Oren Shrout, Stephen Gould:
3DInAction: Understanding Human Actions in 3D Point Clouds. 19978-19987
- Hanfeng Wu, Xingxing Zuo, Stefan Leutenegger, Or Litany, Konrad Schindler, Shengyu Huang:
Dynamic LiDAR Re-Simulation Using Compositional Neural Fields. 19988-19998
- Haoyuan Wang, Wenbo Hu, Lei Zhu, Rynson W. H. Lau:
Inverse Rendering of Glossy Objects via the Neural Plenoptic Function and Radiance Fields. 19999-20008
- Diantao Tu, Hainan Cui, Xianwei Zheng, Shuhan Shen:
PanoPose: Self-supervised Relative Pose Estimation for Panoramic Images. 20009-20018
- Shengjun Zhang, Xin Fei, Yueqi Duan:
GeoAuxNet: Towards Universal 3D Representation Learning for Multi-Sensor Point Clouds. 20019-20028
- Zhen Xu, Sida Peng, Haotong Lin, Guangzhao He, Jiaming Sun, Yujun Shen, Hujun Bao, Xiaowei Zhou:
4K4D: Real-Time 4D View Synthesis at 4K Resolution. 20029-20040
- Haofei Xu, Anpei Chen, Yuedong Chen, Christos Sakaridis, Yulun Zhang, Marc Pollefeys, Andreas Geiger, Fisher Yu:
MuRF: Multi-Baseline Radiance Fields. 20041-20050
- Minghan Qin, Wanhua Li, Jiawei Zhou, Haoqian Wang, Hanspeter Pfister:
LangSplat: 3D Language Gaussian Splatting. 20051-20060
- Lily Goli, Cody Reading, Silvia Sellán, Alec Jacobson, Andrea Tagliasacchi:
Bayes' Rays: Uncertainty Quantification for Neural Radiance Fields. 20061-20070
- Shakiba Kheradmand, Daniel Rebain, Gopal Sharma, Hossam Isack, Abhishek Kar, Andrea Tagliasacchi, Kwang Moo Yi:
Accelerating Neural Field Training via Soft Mining. 20071-20080
- Donggeun Yoon, Donghyeon Cho:
CORE-MPI: Consistency Object Removal with Embedding MultiPlane Image. 20081-20090
- Junjin Xiao, Qing Zhang, Zhan Xu, Wei-Shi Zheng:
NECA: Neural Customizable Human Avatar. 20091-20101
- Xingyi Li, Zhiguo Cao, Yizheng Wu, Kewei Wang, Ke Xian, Zhe Wang, Guosheng Lin:
S-DyRF: Reference-Based Stylized Radiance Fields for Dynamic Scenes. 20102-20112
- Zhenxin Li, Shiyi Lan, José M. Álvarez, Zuxuan Wu:
BEVNeXt: Reviving Dense BEV Frameworks for 3D Object Detection. 20113-20123
- Yujie Xue, Ruihui Li, Fan Wu, Zhuo Tang, Kenli Li, Mingxing Duan:
Bi-SSC: Geometric-Semantic Bidirectional Fusion for Camera-Based 3D Semantic Scene Completion. 20124-20134
- Yunzhong Hou, Stephen Gould, Liang Zheng:
Learning to Select Views for Efficient Multi-View Understanding. 20135-20144
- Dongsu Zhang, Francis Williams, Zan Gojcic, Karsten Kreis, Sanja Fidler, Young Min Kim, Amlan Kar:
Outdoor Scene Extrapolation with Hierarchical Generative Cellular Automata. 20145-20154
- Tianyu Luan, Zhong Li, Lele Chen, Xuan Gong, Lichang Chen, Yi Xu, Junsong Yuan:
Spectrum AUC Difference (SAUCD): Human-Aligned 3D Shape Evaluation. 20155-20164
- Matteo Poggi, Fabio Tosi:
Federated Online Adaptation for Deep Stereo. 20165-20175
- Linzhan Mou, Jun-Kun Chen, Yu-Xiong Wang:
Instruct 4D-to-4D: Editing 4D Scenes as Pseudo-3D Scenes Using 2D Diffusion. 20176-20185
- Yixin Zeng, Zoubin Bi, Mingrui Yin, Xiang Feng, Kun Zhou, Hongzhi Wu:
Real-Time Acquisition and Reconstruction of Dynamic Volumes with Neural Structured Illumination. 20186-20195
- Sunghwan Hong, Jaewoo Jung, Heeseong Shin, Jiaolong Yang, Seungryong Kim, Chong Luo:
Unifying Correspondence, Pose and NeRF for Generalized Pose-Free Novel View Synthesis. 20196-20206
- Jiang Wu, Rui Li, Haofei Xu, Wenxun Zhao, Yu Zhu, Jinqiu Sun, Yanning Zhang:
GoMVS: Geometrically Consistent Cost Aggregation for Multi-View Stereo. 20207-20216



- Yesheng Zhang, Xu Zhao:
MESA: Matching Everything by Segmenting Anything. 20217-20226
- Hakyeong Kim, Andreas Meuleman, Hyeonjoong Jang, James Tompkin, Min H. Kim:
OmniSDF: Scene Reconstruction Using Omnidirectional Signed Distance Functions and Adaptive Binoctrees. 20227-20236
- Haowen Sun, Yueqi Duan, Juncheng Yan, Yifan Liu, Jiwen Lu:
MirageRoom: 3D Scene Segmentation with 2D Pre-Trained Models by Mirage Projection. 20237-20246
- Jiawei Zhang, Jiahe Li, Lei Huang, Xiaohan Yu, Lin Gu, Jin Zheng, Xiao Bai:
Robust Synthetic-to-Real Transfer for Stereo Matching. 20247-20257
- Haoyi Jiang, Tianheng Cheng, Naiyu Gao, Haoyang Zhang, Tianwei Lin, Wenyu Liu, Xinggang Wang:
Sympphonize 3D Semantic Scene Completion with Contextual Instance Queries. 20258-20267
- Weijian Deng, Dylan Campbell, Chunyi Sun, Shubham Kanitkar, Matthew E. Shaffer, Stephen Gould:
Differentiable Neural Surface Refinement for Modeling Transparent Objects. 20268-20277
- Shihua Zhang, Zizhuo Li, Yuan Gao, Jiayi Ma:
DeMatch: Deep Decomposition of Motion Field for Two-View Correspondence Learning. 20278-20287
- Hanxin Zhu, Tianyu He, Xin Li, Bingchen Li, Zhibo Chen:
Is Vanilla MLP in Neural Radiance Field Enough for Few-Shot View Synthesis? 20288-20298
- Shenhan Qian, Tobias Kirschstein, Liam Schoneveld, Davide Davoli, Simon Giebenhain, Matthias Nießner:
GaussianAvatars: Photorealistic Head Avatars with Rigged 3D Gaussians. 20299-20309
- Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, Xinggang Wang:
4D Gaussian Splatting for Real-Time Dynamic Scene Rendering. 20310-20320
- Yihang Chen, Qianyi Wu, Mehrtash Harandi, Jianfei Cai:
How Far can we Compress Instant-NGP-Based NeRF? 20321-20330
- Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, Xiaogang Jin:
Deformable 3D Gaussians for High-Fidelity Monocular Dynamic Scene Reconstruction. 20331-20341
- Yingjie Xu, Bangzhen Liu, Hao Tang, Bailin Deng, Shengfeng He:
Learning with Unreliability: Fast Few-Shot Voxel Radiance Fields with Relative Geometric Consistency. 20342-20351
- Xiaobao Wei, Renrui Zhang, Jiarui Wu, Jiaming Liu, Ming Lu, Yandong Guo, Shanghang Zhang:
NTO3D: Neural Target Object 3D Reconstruction with Segment Anything. 20352-20362
- Lorenzo Liso, Erik Sandström, Vladimir Yugay, Luc Van Gool, Martin R. Oswald:
Loopy-SLAM: Dense Neural SLAM with Loop Closures. 20363-20373
- Jiahao Lu, Jiacheng Deng, Tianzhu Zhang:
BSNet: Box-Supervised Simulation-Assisted Mean Teacher for 3D Instance Segmentation. 20374-20384
- Meng-Li Shih, Wei-Chiu Ma, Lorenzo Boyice, Aleksander Holynski, Forrester Cole, Brian Curless, Janne Kontkanen:
ExtraNeRF: Visibility-Aware View Extrapolation of Neural Radiance Fields with Diffusion Models. 20385-20395
- Joshua Ahn, Haochen Wang, Raymond A. Yeh, Greg Shakhnarovich:
Alpha Invariance: On Inverse Scaling Between Distance and Volume Density in Neural Radiance Fields. 20396-20405
- Yuxi Xiao, Qianqian Wang, Shangzhan Zhang, Nan Xue, Sida Peng, Yujun Shen, Xiaowei Zhou:
SpatialTracker: Tracking Any 2D Pixels in 3D Space. 20406-20417
- Shoukang Hu, Tao Hu, Ziwei Liu:
GauHuman: Articulated Gaussian Splatting from Monocular Human Videos. 20418-20431
- Yushuang Wu, Luyue Shi, Junhao Cai, Weihao Yuan, Lingteng Qiu, Zilong Dong, Liefeng Bo, Shuguang Cui, Xiaoguang Han:
IPoD: Implicit Field Learning with Point Diffusion for Generalizable 3D Object Reconstruction from Single RGB-D Images. 20432-20442
- Yunsong Wang, Hanlin Chen, Gim Hee Lee:
GOV-NeSF: Generalizable Open-Vocabulary Neural Semantic Fields. 20443-20453
- Haolin Liu, Chongjie Ye, Yinyu Nie, Yingfan He, Xiaoguang Han:
LASA: Instance Reconstruction from Real Scans using A Large-scale Aligned Shape Annotation Dataset. 20454-20464
- Lei Li, Angela Dai:
GenZI: Zero-Shot 3D Human-Scene Interaction Generation. 20465-20474
- Hiroaki Santo, Fumio Okura, Yasuyuki Matsushita:
MVCPS-NeuS: Multi-View Constrained Photometric Stereo for Neural Surface Reconstruction. 20475-20484
- Chen Zhao, Tong Zhang, Zheng Dang, Mathieu Salzmann:
DVMNet: Computing Relative Pose for Unseen Objects Beyond Hypotheses. 20485-20495
- Wei Cao, Chang Luo, Biao Zhang, Matthias Nießner, Jiapeng Tang:
Motion2VecSets: 4D Latent Vector Set Diffusion for Non-Rigid Shape Reconstruction and Tracking. 20496-20506
- Jiapeng Tang, Yinyu Nie, Lev Markhasin, Angela Dai, Justus Thies, Matthias Nießner:
DiffuScene: Denoising Diffusion Models for Generative Indoor Scene Synthesis. 20507-20518



- Hyoungseob Park, Anjali Gupta, Alex Wong:
Test- Time Adaptation for Depth Completion. 20519-20529
- Xiaotian Sun, Qingshan Xu, Xinjie Yang, Yu Zang, Cheng Wang:
Global and Hierarchical Geometry Consistency Priors for Few-Shot NeRFs in Indoor Scenes. 20530-20539
- Ruida Zhang, Chenyangguang Zhang, Yan Di, Fabian Manhardt, Xingyu Liu, Federico Tombari, Xiangyang Ji:
KP-RED: Exploiting Semantic Keypoints for Joint 3D Shape Retrieval and Deformation. 20540-20550
- Yujie Lu, Long Wan, Nayu Ding, Yulong Wang, Shuhan Shen, Shen Cai, Lin Gao:
Unsigned Orthogonal Distance Fields: An Accurate Neural Implicit Representation for Diverse 3D Shapes. 20551-20560
- Jie Long Lee, Chen Li, Gim Hee Lee:
DiSR-NeRF: Diffusion-Guided View-Consistent Super-Resolution NeRF. 20561-20570
- Akhmedkhan Ahan Shabanov, Shrisudhan Govindarajan, Cody Reading, Lily Goli, Daniel Rebain, Kwang Moo Yi, Andrea Tagliasacchi:
BANF: Band-Limited Neural Fields for Levels of Detail Reconstruction. 20571-20580
- Xu Cao, Takafumi Taketomi:
SuperNormal: Neural Surface Reconstruction via Multi-View Normal Integration. 20581-20590
- Han Ling, Quansen Sun, Yinghui Sun, Xian Xu, Xingfeng Li:
ADFactory: An Effective Framework for Generalizing Optical Flow With NeRF. 20591-20600
- Yusuke Takimoto, Hikari Takehara, Hiroyuki Sato, Zihao Zhu, Bo Zheng:
Dr.Hair: Reconstructing Scalp-Connected Hair Strands without Pre-Training via Differentiable Rendering of Line Segments. 20601-20611
- Haiyang Ying, Yixuan Yin, Jinzhi Zhang, Fan Wang, Tao Yu, Ruqi Huang, Lu Fang:
OmniSeg3D: Omniversal 3D Segmentation via Hierarchical Contrastive Learning. 20612-20622
- Zhihao Yuan, Jinke Ren, Chun-Mei Feng, Hengshuang Zhao, Shuguang Cui, Zhen Li:
Visual Programming for Zero-Shot Open-Vocabulary 3D Visual Grounding. 20623-20633
- Keyang Zhou, Bharat Lal Bhatnagar, Jan Eric Lenssen, Gerard Pons-Moll:
GEARS: Local Geometry-Aware Hand-Object Interaction Synthesis. 20634-20643
- Wonseok Roh, Hwanhee Jung, Giljoo Nam, Jinseop Yeom, Hyunje Park, Sang Ho Yoon, Sangpil Kim:
Edge-Aware 3D Instance Segmentation Network with Intelligent Semantic Prior. 20644-20653
- Tao Lu, Mulin Yu, Lining Xu, Yuanbo Xiangli, Limin Wang, Dahua Lin, Bo Dai:
Scaffold-GS: Structured 3D Gaussians for View-Adaptive Rendering. 20654-20664
- Shuai Chen, Tommaso Cavallari, Victor Adrian Prisacariu, Eric Brachmann:
Map-Relative Pose Regression for Visual Re-Localization. 20665-20674
- Jiakai Sun, Han Jiao, Guangyuan Li, Zhanjie Zhang, Lei Zhao, Wei Xing:
3DGStream: On-the-Fly Training of 3D Gaussians for Efficient Streaming of Photo-Realistic Free-Viewpoint Videos. 20675-20685
- Peilin Tao, Hainan Cui, Mengqi Rong, Shuhan Shen:
Revisiting Global Translation Estimation with Feature Tracks. 20686-20696
- Shuzhe Wang, Vincent Leroy, Yohann Cabon, Boris Chidlovskii, Jérôme Revaud:
DUSt3R: Geometric 3D Vision Made Easy. 20697-20709
- Kei Ikemura, Yiming Huang, Felix Heide, Zhaoxiang Zhang, Qifeng Chen, Chenyang Lei:
Robust Depth Enhancement via Polarization Prompt Fusion Tuning. 20710-20720
- Dasith de Silva Edirimuni , Xuequan Lu, Gang Li, Lei Wei, Antonio Robles-Kelly, Hongdong Li:
StraightPCF: Straight Point Cloud Filtering. 20721-20730
- Ethan Weber, Aleksander Holynski, Varun Jampani, Saurabh Saxena, Noah Snavely, Abhishek Kar, Angjoo Kanazawa:
NeRFiller: Completing Scenes via Generative 3D Inpainting. 20731-20741
- Wenhui Xiao, Rodrigo Santa Cruz, David Ahmedt-Aristizabal, Olivier Salvado, Clinton Fookes, Léo Lebrat:
NeRF Director: Revisiting View Selection in Neural Volume Rendering. 20742-20751
- Rui Gong, Weide Liu, Zaiwang Gu, Xulei Yang, Jun Cheng:
Learning Intra-View and Cross-View Geometric Knowledge for Stereo Matching. 20752-20762
- Fangfu Liu, Diankun Wu, Yi Wei, Yongming Rao, Yueqi Duan:
Sherpa3D: Boosting High-Fidelity Text-to-3D Generation via Coarse 3D Prior. 20763-20774
- Jiahe Li, Jiawei Zhang, Xiao Bai, Jin Zheng, Xin Ning, Jun Zhou, Lin Gu:
DNGaussian: Optimizing Sparse-View 3D Gaussian Radiance Fields with Global-Local Depth Normalization. 20775-20785
- Wentao Qu, Yuantian Shao, Lingwu Meng, Xiaoshui Huang, Liang Xiao:
A Conditional Denoising Diffusion Probabilistic Model for Point Cloud Upsampling. 20786-20795
- Yang Fu, Xiaolong Wang, Sifei Liu, Amey Kulkarni, Jan Kautz, Alexei A. Efros:
COLMAP-Free 3D Gaussian Splatting. 20796-20805



- Zi-Ting Chou, Sheng-Yu Huang, I-Jieh Liu, Yu-Chiang Frank Wang:
GSNeRF: Generalizable Semantic Neural Radiance Fields with Enhanced 3D Scene Understanding. 20806-20815
- Quan Liu, Hongzi Zhu, Zhenxi Wang, Yunsong Zhou, Shan Chang, Minyi Guo:
Extend Your Own Correspondences: Unsupervised Distant Point Cloud Registration by Progressive Distance Extension. 20816-20826
- Junho Kim, Jiwon Jeong, Young Min Kim:
Fully Geometric Panoramic Localization. 20827-20837
- Shengze Jin, Iro Armeni, Marc Pollefeys, Dániel Baráth:
Multiway Point Cloud Mosaicking with Diffusion and Global Optimization. 20838-20849
- Bi'an Du, Xiang Gao, Wei Hu, Renjie Liao:
Generative 3D Part Assembly via Part-Whole-Hierarchy Message Passing. 20850-20859
- Xiaoyang Lyu, Chirui Chang, Peng Dai, Yang-Tian Sun, Xiaojuan Qi:
Total-Decom: Decomposed 3D Scene Reconstruction with Minimal Interaction. 20860-20869
- Jonathan Ventura, Zuzana Kukelova, Torsten Sattler, Dániel Baráth:
Absolute Pose from One or Two Scaled and Oriented Features. 20870-20880
- Shuzhe Wang, Juho Kannala, Daniel Barath:
DGC-GNN: Leveraging Geometry and Color Cues for Visual Descriptor-Free 2D-3D Matching. 20881-20891
- Takashi Otonari, Satoshi Ikehata, Kiyoharu Aizawa:
Entity-NeRF: Detecting and Removing Moving Entities in Urban Scenes. 20892-20901
- Junjie Wang, Jiemin Fang, Xiaopeng Zhang, Lingxi Xie, Qi Tian:
GaussianEditor: Editing 3D Gaussians Delicately with Text Instructions. 20902-20911
- Xinyang Han, Zelin Gao, Angjoo Kanazawa, Shubham Goel, Yossi Gandelman:
The More You See in 2D, the More You Perceive in 3D. 20912-20922
- Zhiwen Yan, Weng Fei Low, Yu Chen, Gim Hee Lee:
Multi-Scale 3D Gaussian Splatting for Anti-Aliased Rendering. 20923-20931
- Zhenyu Chen, Jie Guo, Shuichang Lai, Ruoyu Fu, Mengxun Kong, Chen Wang, Hongyu Sun, Zhebin Zhang, Chen Li, Yanwen Guo:
Practical Measurements of Translucent Materials with Inter-Pixel Translucency Prior. 20932-20942
- Maxim Kolodiaznyi, Anna Vorontsova, Anton Konushin, Danila Rukhovich:
OneFormer3D: One Transformer for Unified Point Cloud Segmentation. 20943-20953
- Zhe Li, Zhangyang Gao, Cheng Tan, Bocheng Ren, Laurence T. Yang, Stan Z. Li:
General Point Model Pretraining with Autoencoding and Autoregressive. 20954-20964
- Hengyi Wang, Jingwen Wang, Lourdes Agapito:
MorpheuS: Neural Dynamic \$360^\circ\$ Surface Reconstruction from Monocular RGB-D Video. 20965-20976
- Chanho Kim, Fuxin Li:
Object Dynamics Modeling with Hierarchical Point Cloud-Based Representations. 20977-20986
- Shuai Chen, Yash Bhalgat, Xinghui Li, Jia-Wang Bian, Kejie Li, Zirui Wang, Victor Adrian Prisacariu:
Neural Refinement for Absolute Pose Regression with Feature Synthesis. 20987-20996
- Luis Bolanos, Shih-Yang Su, Helge Rhodin:
Gaussian Shadow Casting for Neural Characters. 20997-21006
- Shichong Peng, Yanshu Zhang, Ke Li:
PAPR in Motion: Seamless Point-level 3D Scene Interpolation. 21007-21016
- Yan Di, Chenyangguang Zhang, Chaowei Wang, Ruida Zhang, Guangyao Zhai, Yanyan Li, Bowen Fu, Xiangyang Ji, Shan Gao:
ShapeMatcher: Self-Supervised Joint Shape Canonicalization, Segmentation, Retrieval and Deformation. 21017-21028
- Guangyu Wang, Jinzhi Zhang, Fan Wang, Ruqi Huang, Lu Fang:
XScale-NVS: Cross-Scale Novel View Synthesis with Hash Featurized Manifold. 21029-21039
- Xiao Lin, Wenfei Yang, Yuan Gao, Tianzhu Zhang:
Instance-Adaptive and Geometric-Aware Keypoint Learning for Category-Level 6D Object Pose Estimation. 21040-21049
- Yi Rong, Haoran Zhou, Kang Xia, Cheng Mei, Jiahao Wang, Tong Lu:
RepKPU: Point Cloud Upsampling with Kernel Point Representation and Deformation. 21050-21060
- Juncheng Mu, Lin Bie, Shaoyi Du, Yue Gao:
ColorPCR: Color Point Cloud Registration with Multi-Stage Geometric-Color Fusion. 21061-21070
- Jun-Kun Chen, Samuel Rota Bulò, Norman Müller, Lorenzo Porzi, Peter Kortschieder, Yu-Xiong Wang:
ConsistDreamer: 3D-Consistent 2D Diffusion for High-Fidelity Scene Editing. 21071-21080
- Dave Zhenyu Chen, Haoxuan Li, Hsin-Ying Lee, Sergey Tulyakov, Matthias Nießner:
SceneTex: High-Quality Texture Synthesis for Indoor Scenes via Diffusion Priors. 21081-21091
- Yuqi Zhang, Guanying Chen, Jiaxing Chen, Shuguang Cui:
Aerial Lifting: Neural Urban Semantic and Building Instance Lifting from Aerial Imagery. 21092-21103



- Yufei Wang, Ge Zhang, Shaoqian Wang, Bo Li, Qi Liu, Le Hui, Yuchao Dai:
Improving Depth Completion via Depth Feature Upsampling. 21104-21113
- Ruoxi Shi, Xinyue Wei, Cheng Wang, Hao Su:
ZeroRF: Fast Sparse View 360° Reconstruction with Zero Pretraining. 21114-21124
- Tobias Fischer, Lorenzo Porzi, Samuel Rota Bulò, Marc Pollefeys, Peter Kontschieder:
Multi-Level Neural Scene Graphs for Dynamic Urban Environments. 21125-21135
- Youtian Lin, Zuozhuo Dai, Siyu Zhu, Yao Yao:
Gaussian-Flow: 4D Reconstruction with Dynamic 3D Gaussian Particle. 21136-21145
- Jingtao Sun, Yaonan Wang, Mingtao Feng, Yulan Guo, Ajmal Mian, Mike Zheng Shou:
L4D-Track: Language-to-4D Modeling Towards 6-DoF Tracking and Shape Reconstruction in 3D Point Cloud Stream. 21146-21156
- Liwen Wu, Sai Bi, Zexiang Xu, Fujun Luan, Kai Zhang, Iliyan Georgiev, Kalyan Sunkavalli, Ravi Ramamoorthi:
Neural Directional Encoding for Efficient and Accurate View-Dependent Appearance Modeling. 21157-21166
- Siting Zhu, Guangming Wang, Hermann Blum, Jiuming Liu, Liang Song, Marc Pollefeys, Hesheng Wang:
SNI-SLAM: Semantic Neural Implicit SLAM. 21167-21177
- Haoxuanye Ji, Pengpeng Liang, Erkang Cheng:
Enhancing 3D Object Detection with 2D Detection-Guided Query Anchors. 21178-21187
- Li Ma, Vasu Agrawal, Haithem Turki, Changil Kim, Chen Gao, Pedro V. Sander, Michael Zollhöfer, Christian Richardt:
SpecNeRF: Gaussian Directional Encoding for Specular Reflections. 21188-21198
- Mingyang Zhao, Jingen Jiang, Lei Ma, Shiqing Xin, Gaofeng Meng, Dong-Ming Yan:
Correspondence-Free Non-Rigid Point Set Registration Using Unsupervised Clustering Analysis. 21199-21208
- Xiaotian Li, Baojie Fan, Jiandong Tian, Huijie Fan:
GAFusion: Adaptive Fusing LiDAR and Camera with Multiple Guidance for 3D Object Detection. 21209-21218
- Lei Li, Songyou Peng, Zehao Yu, Shaohui Liu, Rémi Pautrat, Xiaochuan Yin, Marc Pollefeys:
3D Neural Edge Reconstruction. 21219-21229
- Tang Tao, Guangrun Wang, Yixing Lao, Peng Chen, Jie Liu, Liang Lin, Kaicheng Yu, Xiaodan Liang:
AlignMiF: Geometry-Aligned Multimodal Implicit Field for LiDAR-Camera Joint Synthesis. 21230-21240
- Dominik Scheuble, Chenyang Lei, Seung-Hwan Baek, Mario Bijelic, Felix Heide:
Polarization Wavefront Lidar: Learning Large Scene Reconstruction from Polarized Wavefronts. 21241-21250
- Jiangnan Tang, Jingya Wang, Kaiyang Ji, Lan Xu, Jingyi Yu, Ye Shi:
A Unified Diffusion Framework for Scene-aware Human Motion Estimation from Sparse Signals. 21251-21262
- Shivangi Aneja, Justus Thies, Angela Dai, Matthias Nießner:
FaceTalk: Audio-Driven Motion Diffusion for Neural Parametric Head Models. 21263-21273
- Sicheng Li, Hao Li, Yiyi Liao, Lu Yu:
NeRFCODEC: Neural Feature Compression Meets Neural Radiance Fields for Memory-Efficient Scene Representation. 21274-21283
- Li Jiang, Shaoshuai Shi, Bernt Schiele:
Open-Vocabulary 3D Semantic Segmentation with Foundation Models. 21284-21294
- Gege Gao, Weiyang Liu, Anpei Chen, Andreas Geiger, Bernhard Schölkopf:
GraphDreamer: Compositional 3D Scene Synthesis from Scene Graphs. 21295-21304
- Bohao Peng, Xiaoyang Wu, Li Jiang, Yukang Chen, Hengshuang Zhao, Zhuotao Tian, Jiaya Jia:
OA-CNNs: Omni-Adaptive Sparse CNNs for 3D Semantic Segmentation. 21305-21315
- Petr Hruby, Timothy Duff, Marc Pollefeys:
Efficient Solution of Point-Line Absolute Pose. 21316-21325
- Guanlin Shen, Jingwei Huang, Zhihua Hu, Bin Wang:
CN-RMA: Combined Network with Ray Marching Aggregation for 3D Indoor Object Detection from Multi-View Images. 21326-21335
- Hongyu Zhou, Jiahao Shao, Lu Xu, Dongfeng Bai, Weichao Qiu, Bingbing Liu, Yue Wang, Andreas Geiger, Yiyi Liao:
HUGS: Holistic Urban 3D Scene Understanding via Gaussian Splatting. 21336-21345
- Tongyan Hua, Lin Wang:
Benchmarking Implicit Neural Representation and Geometric Rendering in Real-Time RGB-D SLAM. 21346-21356
- Nikhil Varma Keetha, Jay Karhade, Krishna Murthy Jatavallabhula, Gengshan Yang, Sebastian A. Scherer, Deva Ramanan, Jonathon Luiten:
SplaTAM: Splat, Track & Map 3D Gaussians for Dense RGB-D SLAM. 21357-21366
- Mukund Varma T., Peihao Wang, Zhiwen Fan, Zhangyang Wang, Hao Su, Ravi Ramamoorthi:
Lift3D: Zero-Shot Lifting of Any 2D Vision Model to 3D. 21367-21377
- Bo Sun, Thibault Groueix, Chen Song, Qixing Huang, Noam Aigerman:
TutteNet: Injective 3D Deformations by Composition of 2D Mesh Deformations. 21378-21389
- Liangchen Li, Juyong Zhang:
L0-Sampler: An L0Model Guided Volume Sampling for NeRF. 21390-21400



- Zilong Chen, Feng Wang, Yikai Wang, Huaping Liu:
Text-to-3D using Gaussian Splatting. 21401-21412
- Zhihao Zhang, Shengcao Cao, Yu-Xiong Wang:
TAMM: TriAdapter Multi-Modal Learning for 3D Shape Understanding. 21413-21423
- Jiahui Zhang, Fangneng Zhan, Muyu Xu, Shijian Lu, Eric P. Xing:
FreGS: 3D Gaussian Splatting with Progressive Frequency Regularization. 21424-21433
- Chenhao Li, Taishi Ono, Takeshi Uemori, Hajime Mihara, Alexander Gatto, Hajime Nagahara, Yusuke Moriuchi:
NeISF: Neural Incident Stokes Field for Geometry and Material Estimation. 21434-21445
- Jiawei Shi, Hui Deng, Yuchao Dai:
Non-rigid Structure-from-Motion: Temporally-smooth Procrustean Alignment and Spatially-variant Deformation Modeling. 21446-21455
- Chamin Hewa Koneputugodage, Yizhak Ben-Shabat, Dylan Campbell, Stephen Gould:
Small Steps and Level Sets: Fitting Neural Surface Models with Point Guidance. 21456-21465
- Yingji Zhong, Lanqing Hong, Zhenguo Li, Dan Xu:
CVT-xRF: Contrastive In-Voxel Transformer for 3D Consistent Radiance Fields from Sparse Inputs. 21466-21475
- Yiwen Chen, Zilong Chen, Chi Zhang, Feng Wang, Xiaofeng Yang, Yikai Wang, Zhongang Cai, Lei Yang, Huaping Liu, Guosheng Lin:
GaussianEditor: Swift and Controllable 3D Editing with Gaussian Splatting. 21476-21485
- Junyi Ma, Xieyuanli Chen, Jiawei Huang, Jingyi Xu, Zhen Luo, Jintao Xu, Weihao Gu, Rui Ai, Hesheng Wang:
Cam4DOcc: Benchmark for Camera-Only 4D Occupancy Forecasting in Autonomous Driving Applications. 21486-21495
- Junsheng Zhou, Weiqi Zhang, Baorui Ma, Kanle Shi, Yu-Shen Liu, Zhizhong Han:
UDiFF: Generating Conditional Unsigned Distance Fields with Optimal Wavelet Diffusion. 21496-21506
- Dong Wu, Zike Yan, Hongbin Zha:
PanoRecon: Real-Time Panoptic 3D Reconstruction from Monocular Video. 21507-21518
- Gilles Puy, Spyros Gidaris, Alexandre Boulch, Oriane Siméoni, Corentin Sautier, Patrick Pérez, Andrei Bursuc, Renaud Marlet:
Three Pillars Improving Vision Foundation Model Distillation for Lidar. 21519-21529
- Chung Min Kim, Mingxuan Wu, Justin Kerr, Ken Goldberg, Matthew Tancik, Angjoo Kanazawa:
GARField: Group Anything with Radiance Fields. 21530-21539
- Jinhyung Park, Yu-Jhe Li, Kris Kitani:
Flexible Depth Completion for Sparse and Varying Point Densities. 21540-21550
- Rundi Wu, Ben Mildenhall, Philipp Henzler, Keunhong Park, Ruiqi Gao, Daniel Watson, Pratul P. Srinivasan, Dor Verbin, Jonathan T. Barron, Ben Poole, Aleksander Holynski:
ReconFusion: 3D Reconstruction with Diffusion Priors. 21551-21561
- Ziyue Feng, Huangying Zhan, Zheng Chen, Qingan Yan, Xiangyu Xu, Changjiang Cai, Bing Li, Qilun Zhu, Yi Xu:
NARUTO: Neural Active Reconstruction from Uncertain Target Observations. 21572-21583
- Huajian Huang, Longwei Li, Hui Cheng, Sai-Kit Yeung:
Photo-SLAM: Real-Time Simultaneous Localization and Photorealistic Mapping for Monocular, Stereo, and RGB-D Cameras. 21584-21593
- Xingyi He, Jiaming Sun, Yifan Wang, Sida Peng, Qixing Huang, Hujun Bao, Xiaowei Zhou:
Detector-Free Structure from Motion. 21594-21603
- Xiuwei Xu, Chong Xia, Ziwei Wang, Linqing Zhao, Yueqi Duan, Jie Zhou, Jiwen Lu:
Memory-based Adapters for Online 3D Scene Perception. 21604-21613
- Lizhe Liu, Bohua Wang, Hongwei Xie, Daqi Liu, Li Liu, Zhiqiang Tian, Kuiyuan Yang, Bing Wang:
SurroundSDF: Implicit 3D Scene Understanding Based on Signed Distance Field. 21614-21623
- Heng Yu, Joel Julin, Zoltán Ádám Milacski, Koichiro Niinuma, László A. Jeni:
CoGS: Controllable Gaussian Splatting. 21624-21633
- Xiaoyu Zhou, Zhiwei Lin, Xiaojun Shan, Yongtao Wang, Deqing Sun, Ming-Hsuan Yang:
DrivingGaussian: Composite Gaussian Splatting for Surrounding Dynamic Autonomous Driving Scenes. 21634-21643
- Zhihao Liang, Qi Zhang, Ying Feng, Ying Shan, Kui Jia:
GS-IR: 3D Gaussian Splatting for Inverse Rendering. 21644-21653
- Samuel Brucker, Stefanie Walz, Mario Bijelic, Felix Heide:
Cross-spectral Gated-RGB Stereo Depth Estimation. 21654-21665
- Yifan Wang, Xingyi He, Sida Peng, Dongli Tan, Xiaowei Zhou:
Efficient LoFTR: Semi-Dense Local Feature Matching with Sparse-Like Speed. 21666-21675
- Shijie Zhou, Haoran Chang, Sicheng Jiang, Zhiwen Fan, Zehao Zhu, Dejia Xu, Pradyumna Chari, Suya You, Zhangyang Wang, Achuta Kadambi:
Feature 3DGs: Supercharging 3D Gaussian Splatting to Enable Distilled Feature Fields. 21676-21685



- Jianyuan Wang, Nikita Karaev, Christian Rupprecht, David Novotný:
VGGsFm: Visual Geometry Grounded Deep Structure from Motion. 21686-21697
- Hong Chen, Pei Yan, Sihe Xiang, Yihua Tan:
Dynamic Cues-Assisted Transformer for Robust Point Cloud Registration. 21698-21707
- Hao Li, Dingwen Zhang, Yalun Dai, Nian Liu, Lechao Cheng, Jingfeng Li, Jingdong Wang, Junwei Han:
GP-NeRF: Generalized Perception NeRF for Context-Aware 3D Scene Understanding. 21708-21718
- Joo Chan Lee, Daniel Rho, Xiangyu Sun, Jong Hwan Ko, Eunbyung Park:
Compact 3D Gaussian Representation for Radiance Field. 21719-21728
- Amine Ouasfi, Adnane Boukhayma:
Unsupervised Occupancy Learning from Sparse Point Cloud. 21729-21739
- Yun Liu, Haolin Yang, Xu Si, Ling Liu, Zipeng Li, Yuxiang Zhang, Yebin Liu, Li Yi:
TACO: Benchmarking Generalizable Bimanual Tool-Action-Object Understanding. 21740-21751
- Chenshuang Zhang, Fei Pan, Junmo Kim, In So Kweon, Chengzhi Mao:
ImageNet-D: Benchmarking Neural Network Robustness on Diffusion Synthetic Object. 21752-21762
- Yiming Xie, Henglu Wei, Zhenyi Liu, Xiaoyu Wang, Xiangyang Ji:
SynFog: A Photorealistic Synthetic Fog Dataset Based on End-to-End Imaging Simulation for Advancing Real-World Defogging in Autonomous Driving. 21763-21772
- Jinglin Xu, Guohao Zhao, Sibo Yin, Wenhao Zhou, Yuxin Peng:
FineSports: A Multi-Person Hierarchical Sports Video Dataset for Fine-Grained Action Understanding. 21773-21782
- Alexander Raistrick, Lingjie Mei, Karhan Kayan, David Yan, Yiming Zuo, Beining Han, Hongyu Wen, Meenal Parakh, Stamatios Alexandropoulos, Lahav Lipson, Zeyu Ma, Jia Deng:
Infinigen Indoors: Photorealistic Indoor Scenes using Procedural Generation. 21783-21794
- Mohamed El Banani, Amit Raj, Kevis-Kokitsi Maninis, Abhishek Kar, Yuanzhen Li, Michael Rubinstein, Deqing Sun, Leonidas J. Guibas, Justin Johnson, Varun Jampani:
Probing the 3D Awareness of Visual Foundation Models. 21795-21806
- Ziqi Huang, Yinan He, Jiashuo Yu, Fan Zhang, Chenyang Si, Yuming Jiang, Yuanhan Zhang, Tianxing Wu, Qingyang Jin, Nattapol Chanpaisit, Yaohui Wang, Xinyuan Chen, Limin Wang, Dahua Lin, Yu Qiao, Ziwei Liu:
VBench: Comprehensive Benchmark Suite for Video Generative Models. 21807-21818
- Xu Cao, Tong Zhou, Yunsheng Ma, Wenqian Ye, Can Cui, Kun Tang, Zhipeng Cao, Kaizhao Liang, Ziran Wang, James M. Rehg, Chao Zheng:
MAPLM: A Real-World Large-Scale Vision-Language Benchmark for Map and Traffic Scene Understanding. 21819-21830
- Mingfei Han, Linjie Yang, Xiaojie Jin, Jiashi Feng, Xiaojun Chang, Heng Wang:
Video Recognition in Portrait Mode. 21831-21841
- He Zhang, Shenghao Ren, Haolei Yuan, Jianhui Zhao, Fan Li, Shuangpeng Sun, Zhenghao Liang, Tao Yu, Qiu Shen, Xun Cao:
MMVP: A Multimodal MoCap Dataset with Vision and Pressure Sensors. 21842-21852
- Letian Zhang, Xiaotong Zhai, Zhongkai Zhao, Yongshuo Zong, Xin Wen, Bingchen Zhao:
What If the TV was off? Examining Counterfactual Reasoning Abilities of Multi-modal Language Models. 21853-21862
- Xueqing Deng, Qihang Yu, Peng Wang, Xiaohui Shen, Liang-Chieh Chen:
COCONut: Modernizing COCO Segmentation. 21863-21873
- Peng-Tao Jiang, Yuqi Yang, Yang Cao, Qibin Hou, Ming-Ming Cheng, Chunhua Shen:
Traffic Scene Parsing Through the TSP6K Dataset. 21874-21885
- Ziyang Chen, Israel D. Gebru, Christian Richardt, Anurag Kumar, William Laney, Andrew Owens, Alexander Richard:
Real Acoustic Fields: An Audio-Visual Room Acoustics Dataset and Benchmark. 21886-21896
- Han Yu, Xingxuan Zhang, Renzhe Xu, Jiashuo Liu, Yue He, Peng Cui:
Rethinking the Evaluation Protocol of Domain Generalization. 21897-21908
- Jielin Qiu, Jiacheng Zhu, William Han, Aditesh Kumar, Karthik Mittal, Claire Jin, Zhengyuan Yang, Linjie Li, Jianfeng Wang, Ding Zhao, Bo Li, Lijuan Wang:
MMSum: A Dataset for Multimodal Summarization and Thumbnail Generation of Videos. 21909-21921
- Che-Jui Chang, Danrui Li, Deep Patel, Parth Goel, Honglu Zhou, Seonghyeon Moon, Samuel S. Sohn, Sejong Yoon, Vladimir Pavlovic, Mubbasis Kapadia:
Learning from Synthetic Human Group Activities. 21922-21932
- Yunhan Zhao, Haoyu Ma, Shu Kong, Charless C. Fowlkes:
Instance Tracking in 3D Scenes from Egocentric Videos. 21933-21944
- Hoang-Quan Nguyen, Thanh-Dat Truong, Xuan-Bac Nguyen, Ashley Dowling, Xin Li, Khoa Luu:
Insect-Foundation: A Foundation Model and Large-Scale 1M Dataset for Visual Insect Understanding. 21945-21955
- Yunhua Zhang, Hazel Doughty, Cees G. M. Snoek:
Low-Resource Vision Challenges for Foundation Models. 21956-21966
- Guillaume Astruc, Nicolas Dufour, Ioannis Siglidis, Constantin Aronsson, Nacim Bouia, Stephanie Fu, Romain Loiseau, Van Nguyen Nguyen, Charles Raude, Elliot Vincent, Lintao Xu, Hongyu Zhou, Loïc Landrieu:
OpenStreetView-5M: The Many Roads to Global Visual Geolocation. 21967-21977



- Jiong Wang, Fengyu Yang, Bingliang Li, Wenbo Gou, Danqi Yan, Ailing Zeng, Yijun Gao, Junle Wang, Yanqing Jing, Ruimao Zhang:
FreeMan: Towards Benchmarking 3D Human Pose Estimation Under Real-World Conditions. 21978-21988
- Yanwen Guo, Yuanqi Li, Dayong Ren, Xiaohong Zhang, Jiawei Li, Liang Pu, Changfeng Ma, Xiaoyu Zhan, Jie Guo, Mingqiang Wei, Yan Zhang, Piaopiao Yu, Shuangyu Yang, Donghao Ji, Huisheng Ye, Hao Sun, Yansong Liu, Yinuo Chen, Jiaqi Zhu, Hongyu Liu:
LiDAR-Net: A Real-Scanned 3D Point Cloud Dataset for Indoor Scenes. 21989-21999
- Quan Zhang, Lei Wang, Vishal M. Patel, Xiaohua Xie, Jianhuang Lai:
View-decoupled Transformer for Person Re-identification under Aerial-ground Camera Network. 22000-22009
- Jialong Zuo, Hanyu Zhou, Ying Nie, Feng Zhang, Tianyu Guo, Nong Sang, Yunhe Wang, Changxin Gao:
UFineBench: Towards Text-based Person Retrieval with Ultra-fine Granularity. 22010-22019
- Xiaoqi Zhao, Youwei Pang, Zhenyu Chen, Qian Yu, Lihe Zhang, Hanqi Liu, Jiaming Zuo, Huchuan Lu:
Towards Automatic Power Battery Detection: New Challenge, Benchmark Dataset and Baseline. 22020-22029
- Jianwu Fang, Lei-lei Li, Junfei Zhou, Junbin Xiao, Hongkai Yu, Chen Lv, Jianru Xue, Tat-Seng Chua:
Abductive Ego-View Accident Video Understanding for Safe Driving Perception. 22030-22040
- Yiming Li, Zhiheng Li, Nuo Chen, Moonjun Gong, Zonglin Lyu, Zehong Wang, Peili Jiang, Chen Feng:
Multiagent Multitraversal Multimodal Self-Driving: Open MARS Dataset. 22041-22051
- Tongtong Yuan, Xuange Zhang, Kun Liu, Bo Liu, Chen Chen, Jian Jin, Zhenzhen Jiao:
Towards Surveillance Video-and-Language Understanding: New Dataset, Baselines, and Challenges. 22052-22061
- Benjamin Naoto Chiche, Yuto Horikawa, Ryo Fujita:
Pre-Training Vision Models with Mandelbulb Variations. 22062-22071
- Yifei Huang, Guo Chen, Jilan Xu, Mingfang Zhang, Lijin Yang, Baoqi Pei, Hongjie Zhang, Lu Dong, Yali Wang, Limin Wang, Yu Qiao:
EgoExoLearn: A Dataset for Bridging Asynchronous Ego- and Exo-centric View of Procedural Activities in Real World. 22072-22086
- Simindokht Jahangard, Zhixi Cai, Shiki Wen, Hamid Rezatofighi:
JRDB-Social: A Multifaceted Robotic Dataset for Understanding of Context and Dynamics of Human Interactions Within Social Groups. 22087-22097
- Yujin Jeon, Eunsue Choi, Youngchan Kim, Yunseong Moon, Khalid Omer, Felix Heide, Seung-Hwan Baek:
Spectral and Polarization Vision: Spectro-polarimetric Real-world Dataset. 22098-22108
- Giuseppe Vecchio, Valentin Deschaintre:
MatSynth: A Modern PBR Materials Dataset. 22109-22118
- M. Tao, Bing Bai, Haozhe Lin, Heyuan Wang, Yu Wang, Lin Luo, Lu Fang:
When Visual Grounding Meets Gigapixel-Level Large-Scale Scenes: Benchmark and Approach. 22119-22128
- Cong Ma, Lei Qiao, Chengkai Zhu, Kai Liu, Zelong Kong, Qing Li, Xueqi Zhou, Yuheng Kan, Wei Wu:
HoloVic: Large-scale Dataset and Benchmark for Multi-Sensor Holographic Intersection and Vehicle-Infrastructure Cooperative. 22129-22138
- Yaofang Liu, Xiaodong Cun, Xuebo Liu, Xintao Wang, Yong Zhang, Haoxin Chen, Yang Liu, Tieyong Zeng, Raymond Chan, Ying Shan:
EvalCrafter: Benchmarking and Evaluating Large Video Generation Models. 22139-22149
- Adam Lilja, Junsheng Fu, Erik Stenborg, Lars Hammarstrand:
Localization is All You Evaluate: Data Leakage in Online Mapping Datasets and How to Fix it. 22150-22159
- Lu Ling, Yichen Sheng, Zhi Tu, Wentian Zhao, Cheng Xin, Kun Wan, Lantao Yu, Qianyu Guo, Zixun Yu, Yawen Lu, Xuanmao Li, Xingpeng Sun, Rohan Ashok, Aniruddha Mukherjee, Hao Kang, Xiangrui Kong, Gang Hua, Tianyi Zhang, Bedrich Benes, Aniket Bera:
DL3DV-10K: A Large-Scale Scene Dataset for Deep Learning-based 3D Vision. 22160-22169
- Yutao Hu, Tianbin Li, Quanfeng Lu, Wenqi Shao, Junjun He, Yu Qiao, Ping Luo:
OmniMedVQA: A New Large-Scale Comprehensive Evaluation Benchmark for Medical LVLM. 22170-22183
- Paul Gavrikov, Janis Keuper:
Can Biases in ImageNet Models Explain Generalization? 22184-22194
- Kunchang Li, Yali Wang, Yinan He, Yizhuo Li, Yi Wang, Yi Liu, Zun Wang, Jilan Xu, Guo Chen, Ping Lou, Limin Wang, Yu Qiao:
MVBench: A Comprehensive Multi-modal Video Understanding Benchmark. 22195-22206
- Wenqiao Li, Xiaohao Xu, Yao Gu, Bozhong Zheng, Shenghua Gao, Yingna Wu:
Towards Scalable 3D Anomaly Detection and Localization: A Benchmark via 3D Anomaly Synthesis and A Self-Supervised Learning Network. 22207-22216
- Sabarinath Mahadevan, Idil Esen Zulfikar, Paul Voigtlaender, Bastian Leibe:
Point-VOS: Pointing Up Video Object Segmentation. 22217-22226
- Tong Wu, Guandao Yang, Zhibing Li, Kai Zhang, Ziwei Liu, Leonidas J. Guibas, Dahua Lin, Gordon Wetzstein:
GPT-4V(ision) is a Human-Aligned Evaluator for Text-to-3D Generation. 22227-22238
- Andrea Rosasco, Stefano Berti, Giulia Pasquale, Damiano Malafronte, Shogo Sato, Hiroyuki Segawa, Tetsugo Inada, Lorenzo Natale:
ConCon-Chi: Concept-Context Chimera Benchmark for Personalized Vision-Language Tasks. 22239-22248



- Lisa Mais, Peter Hirsch, Claire Managan, Ramya Kandarpa, Josef Lorenz Rumberger, Annika Reinke, Lena Maier-Hein, Gudrun Ihrke, Dagmar Kainmueller:
FISBe: A Real-World Benchmark Dataset for Instance Segmentation of Long-Range thin Filamentous Structures. 22249-22259
- Liang Xu, Xintao Lv, Yichao Yan, Xin Jin, Shuwen Wu, Congsheng Xu, Yifan Liu, Yizhou Zhou, Fengyun Rao, Xingdong Sheng, Yunhui Liu, Wenjun Zeng, Xiaokang Yang:
Inter-X: Towards Versatile Human-Human Interaction Analysis. 22260-22271
- Jialei Cui, Jianwei Du, Wenzhuo Liu, Zhouhui Lian:
TextNeRF: A Novel Scene-Text Image Synthesis Method Based on Neural Radiance Fields. 22272-22281
- Zhe Huang, Ruijie Jiang, Shuchin Aeron, Michael C. Hughes:
Systematic comparison of semi-supervised and self-supervised learning for medical image classification. 22282-22293
- Eunsu Baek, Keondo Park, Jiyo Kim, Hyung-Sin Kim:
Unexplored Faces of Robustness and Out-of-Distribution: Covariate Shifts in Environment and Sensor Domains. 22294-22303
- Thien-Minh Nguyen, Shenghai Yuan, Thien Hoang Nguyen, Pengyu Yin, Haozhi Cao, Lihua Xie, Maciej K. Wozniak, Patric Jensfelt, Marko Thiel, Justin Zieggenbein, Noel Blunder:
MCD: Diverse Large-Scale Multi-Campus Dataset for Robot Perception. 22304-22313
- Huajian Huang, Changkun Liu, Yipeng Zhu, Hui Cheng, Tristan Braud, Sai-Kit Yeung:
360Loc: A Dataset and Benchmark for Omnidirectional Visual Localization with Cross-Device Queries. 22314-22324
- Duy-Tho Le, Chenhui Gou, Stavya Datta, Hengcan Shi, Ian D. Reid, Jianfei Cai, Hamid Rezatofighi:
JRDB-PanoTrack: An Open-World Panoptic Segmentation and Tracking Robotic Dataset in Crowded Human Environments. 22325-22334
- Sanghyun Woo, Kwanyong Park, Inkyu Shin, Myungchul Kim, In So Kweon:
MTMMC: A Large-Scale Real-World Multi-Modal Camera Tracking Benchmark. 22335-22346
- Ruiyang Hao, Siqi Fan, Yingru Dai, Zhenlin Zhang, Chenxi Li, Yuntian Wang, Haibao Yu, Wenxian Yang, Jirui Yuan, Zaiqing Nie:
RCooper: A Real-world Large-scale Dataset for Roadside Cooperative Perception. 22347-22357
- Yaofeng Xie, Lingwei Kong, Kai Chen, Ziqiang Zheng, Xiao Yu, Zhibin Yu, Bing Zheng:
UVEB: A Large-scale Benchmark and Baseline Towards Real-World Underwater Video Enhancement. 22358-22367
- Roman Flepp, Andrey Ignatov, Radu Timofte, Luc Van Gool:
Real-World Mobile Image Denoising Dataset with Efficient Baselines. 22368-22377
- Hongchi Xia, Yang Fu, Sifei Liu, Xiaolong Wang:
RGBD Objects in the Wild: Scaling Real-World 3D Object Learning from RGB-D Videos. 22378-22389
- Mengyu Dai, Amir Hossein Raffiee, Aashish Jain, Joshua Correa:
Evaluating Transferability in Retrieval Tasks: An Approach Using MMD and Kernel Methods. 22390-22400
- Yunhao Ge, Yihe Tang, Jiashu Xu, Cem Gokmen, Chengshu Li, Wensi Ai, Benjamin Jose Martinez, Arman Aydin, Mona Anvari, Ayush K. Chakravarthy, Hong-Xing Yu, Josiah Wong, Sanjana Srivastava, Sharon Lee, Shengxin Zha, Laurent Itti, Yunzhu Li, Roberto Martín-Martín, Miao Liu, Pengchuan Zhang, Ruohan Zhang, Li Fei-Fei, Jiajun Wu:
BEHAVIOR Vision Suite: Customizable Dataset Generation via Simulation. 22401-22412
- Yu-Bang Zheng, Xi-Le Zhao, Junhua Zeng, Chao Li, Qibin Zhao, Heng-Chao Li, Ting-Zhu Huang:
SVDinsTN: A Tensor Network Paradigm for Efficient Structure Search from Regularized Modeling Perspective. 22413-22422
- Petru-Daniel Tudosiu, Yongxin Yang, Shifeng Zhang, Fei Chen, Steven McDonagh, Gerasimos Lampouras, Ignacio Iacobacci, Sarah Parisot:
MULAN: A Multi Layer Annotated Dataset for Controllable Text-to-Image Generation. 22413-22422
- Anas Mahmoud, Mostafa Elhoushi, Amro Abbas, Yu Yang, Newsha Ardalani, Hugh Leather, Ari S. Morcos:
Sieve: Multimodal Dataset Pruning Using Image Captioning Models. 22423-22432
- Peibei Cao, Rafal K. Mantiuk, Kede Ma:
Perceptual Assessment and Optimization of HDR Image Rendering. 22433-22443
- Mohammad Reza Taesiri, Tianjun Feng, Cor-Paul Bezemer, Anh Nguyen:
GlitchBench: Can Large Multimodal Models Detect Video Game Glitches? 22444-22455
- Tom Kelly, John Femiani, Peter Wonka:
WinSyn: A High Resolution Testbed for Synthetic Data. 22456-22465
- Cheng-You Lu, Peisen Zhou, Angela Xing, Chandradeep Pokhariya, Arnab Dey, Ishaan Nikhil Shah, Rugved Mavidipalli, Dylan Hu, Andrew I. Comport, Kefan Chen, Srinath Sridhar:
DiVa-360: The Dynamic Visual Dataset for Immersive Neural Fields. 22466-22476
- Suyeon Kim, Dongha Lee, SeongKu Kang, Sukang Chae, Sanghwan Jang, Hwanjo Yu:
Learning Discriminative Dynamics with Label Corruption for Noisy Label Detection. 22477-22487
- Arjun Balasingam, Joseph Chandler, Chenning Li, Zhoutong Zhang, Hari Balakrishnan:
DriveTrack: A Benchmark for Long-Range Point Tracking in Real-World Videos. 22488-22497



- HyunJun Jung, Shun-Cheng Wu, Patrick Ruhkamp, Guangyao Zhai, Hannah Schieber, Giulia Rizzoli, Pengyuan Wang, Hongcheng Zhao, Lorenzo Garattoni, Daniel Roth, Sven Meier, Nassir Navab, Benjamin Busam: **HouseCat6D - A Large-Scale Multi-Modal Category Level 6D Object Perception Dataset with Household Objects in Realistic Scenarios.** 22498-22508
- Zijin Yin, Kongming Liang, Bing Li, Zhanyu Ma, Jun Guo: **Benchmarking Segmentation Models with Mask-Preserved Attribute Editing.** 22509-22519
- Lorenzo Bianchi, Fabio Carrara, Nicola Messina, Claudio Gennaro, Fabrizio Falchi: **The Devil is in the Fine-Grained Details: Evaluating open-Vocabulary Object Detectors for Fine-Grained Understanding.** 22520-22529
- Xiaoyun Zheng, Liwei Liao, Xufeng Li, Jianbo Jiao, Rongjie Wang, Feng Gao, Shiqi Wang, Ronggang Wang: **PKU-DyMVHumans: A Multi-View Video Benchmark for High-Fidelity Dynamic Human Modeling.** 22530-22540
- Rob Geda, David Towers, Matthew Forshaw, Amir Atapour-Abarghouei, A. Stephen McGough: **Insights from the Use of Previously Unseen Neural Architecture Search Datasets.** 22541-22550
- Kyungdo Kim, Sihan Lyu, Sneha Mantri, Timothy W. Dunn: **TULIP: Multi-Camera 3D Precision Assessment of Parkinson's Disease.** 22551-22562
- Jing Zhang, Irving Fang, Hao Wu, Akshat Kaushik, Alice Rodriguez, Hanwen Zhao, Juexiao Zhang, Zhuo Zheng, Radu Iovita, Chen Feng: **LUWA Dataset: Learning Lithic Use-Wear Analysis on Microscopic Images.** 22563-22573
- Habib Slim, Mohamed Elhoseiny: **ShapeWalk: Compositional Shape Editing Through Language-Guided Chains.** 22574-22583
- Ruiyi Zhang, Yanzhe Zhang, Jian Chen, Yufan Zhou, Jiuxiang Gu, Changyou Chen, Tong Sun: **TRINS: Towards Multimodal Language Models that Can Read.** 22584-22594
- Ryan D. Burgert, Brian L. Price, Jason Kuen, Yijun Li, Michael S. Ryoo: **MAGICK: A Large-Scale Captioned Dataset from Matting Generated Images Using Chroma Keying.** 22595-22604
- Trung Tuan Dao, Duc Hong Vu, Cuong Pham, Anh Tuan Tran: **EFHQ: Multi-Purpose ExtremePose-Face-HQ Dataset.** 22605-22615
- Samuele Papa, Riccardo Valperga, David M. Knigge, Miltiadis Kofinas, Phillip Lippe, Jan-Jakob Sonke, Efstratios Gavves: **How to Train Neural Field Representations: A Comprehensive Study and Benchmark.** 22616-22625
- Galadrielle Humbot-Renaux , Sergio Escalera , Thomas B. Moeslund : **A Noisy Elephant in the Room: Is Your out-of-Distribution Detector Robust to Label Noise?** 22626-22636
- Aayush Atul Verma, Bharatesh Chakravarthi, Arpitsinh Vaghela, Hua Wei, Yezhou Yang: **eTraM: Event-Based Traffic Monitoring Dataset.** 22637-22646
- Shibo Zhao, Yuanjun Gao, Tianhao Wu, Damanpreet Singh, Rushan Jiang, Haoxiang Sun, Mansi Sarawata, Yuheng Qiu, Warren Whittaker, Ian Higgins, Yi Du, Shaoshu Su, Can Xu, John Keller, Jay Karhade, Lucas Nogueira, Sourojit Saha, Ji Zhang, Wenshan Wang, Chen Wang, Sebastian A. Scherer: **SubT-MRS Dataset: Pushing SLAM Towards All-weather Environments.** 22647-22657
- Daniel Kent, Mohammed Alyaqoub, Xiaohu Lu, Hamed Khatounabadi, Kookjin Sung, Cole Scheller, Alexander Dalat, Xinwei Guo, Asma bin Thabit, Roberto Whitley, Hayder Radha: **MSU-4S - The Michigan State University Four Seasons Dataset.** 22658-22667
- Walter Zimmer, Gerhard Arya Wardana, Suren Sritharan, Xingcheng Zhou, Rui Song, Alois C. Knoll: **TUMTraf V2X Cooperative Perception Dataset.** 22668-22677
- Aritra Dutta, Srijan Das, Jacob Nielsen, Rajatsubhra Chakraborty, Mubarak Shah: **Multiview Aerial Visual Recognition (MAVREC): Can Multi-View Improve Aerial Visual Perception?** 22678-22690
- Agastya Kalra, Guy Stoppi, Dmitrii Marin, Vage Taamazyan, Aarrushi Shandilya, Rishav Agarwal, Anton Boykov, Tze Hao Chong, Michael Stark: **Towards Co-Evaluation of Cameras, HDR, and Algorithms for Industrial-Grade 6DoF Pose Estimation.** 22691-22701
- Sachin Goyal, Pratyush Maini, Zachary C. Lipton, Aditi Raghunathan, J. Zico Kolter: **Scaling Laws for Data Filtering - Data Curation Cannot be Compute Agnostic.** 22702-22711
- Chen Liu, Peike Patrick Li, Qingtao Yu, Hongwei Sheng, Dadong Wang, Lincheng Li, Xin Yu: **Benchmarking Audio Visual Segmentation for Long-Untrimmed Videos.** 22712-22722
- Yeonguk Yu, Sungho Shin, Seunghyeok Back, Minhwan Ko, Sangjun Noh, Kyoobin Lee: **Domain-Specific Block Selection and Paired-View Pseudo-Labeling for Online Test-Time Adaptation.** 22723-22732
- Gensheng Pei, Tao Chen, Xiruo Jiang, Huafeng Liu, Zeren Sun, Yazhou Yao: **VideoMAC: Video Masked Autoencoders Meet ConvNets.** 22733-22743
- Dantong Niu, Xudong Wang, Xinyang Han, Long Lian, Roei Herzig, Trevor Darrell: **Unsupervised Universal Image Segmentation.** 22744-22754
- Xudong Wang, Ishan Misra, Ziyun Zeng, Rohit Girdhar, Trevor Darrell: **VideoCutLER: Surprisingly Simple Unsupervised Video Instance Segmentation.** 22755-22764
- Alex Trevithick, Matthew A. Chan, Towaki Takikawa, Umar Iqbal, Shalini De Mello, Manmohan Chandraker, Ravi Ramamoorthi, Koki Nagano: **What You See is What You GAN: Rendering Every Pixel for High-Fidelity Geometry in 3D GANs.** 22765-22775



- Ioannis Kakogeorgiou, Spyros Gidaris, Konstantinos Karantzalos, Nikos Komodakis:
SPOT: Self-Training with Patch-Order Permutation for Object-Centric Learning with Autoregressive Transformers. 22776-22786
- Leonhard Sommer, Artur Jesslen, Eddy Ilg, Adam Kortylewski:
Unsupervised Learning of Category-Level 3D Pose from Object-Centric Videos. 22787-22796
- Fengda Zhang, Qianpei He, Kun Kuang, Jiashuo Liu, Long Chen, Chao Wu, Jun Xiao, Hanwang Zhang:
Distributionally Generative Augmentation for Fair Facial Attribute Classification. 22797-22808
- Rui Zhao, Bin Shi, Jianfei Ruan, Tianze Pan, Bo Dong:
Estimating Noisy Class Posterior with Part-level Labels for Noisy Label Learning. 22809-22819
- Eric Hedlin, Gopal Sharma, Shweta Mahajan, Xingzhe He, Hossam Isack, Abhishek Kar, Helge Rhodin, Andrea Tagliasacchi, Kwang Moo Yi:
Unsupervised Keypoints from Pretrained Diffusion Models. 22820-22830
- Yang Luo, Zhineng Chen, Peng Zhou, Zuxuan Wu, Xieping Gao, Yu-Gang Jiang:
Learning to Rank Patches for Unbiased Image Redundancy Reduction. 22831-22840
- Xinting Liao, Weiming Liu, Chaochao Chen, Pengyang Zhou, Fengyuan Yu, Huabin Zhu, Binhui Yao, Tao Wang, Xiaolin Zheng, Yanchao Tan:
Rethinking the Representation in Federated Unsupervised Learning with Non-IID Data. 22841-22850
- Jihao Liu, Jinliang Zheng, Yu Liu, Hongsheng Li:
GLID: Pre-training a Generalist Encoder-Decoder Vision Model. 22851-22860
- Yutong Bai, Xinyang Geng, Karttikeya Mangalam, Amir Bar, Alan L. Yuille, Trevor Darrell, Jitendra Malik, Alexei A. Efros:
Sequential Modeling Enables Scalable Learning for Large Vision Models. 22861-22872
- Linshan Wu, Jiaxin Zhuang, Hao Chen:
VoCo: A Simple-Yet-Effective Volume Contrastive Learning Framework for 3D Medical Image Analysis. 22873-22882
- Chengjie Wang, Wenbing Zhu, Bin-Bin Gao, Zhenye Gan, Jiangning Zhang, Zhihao Gu, Shuguang Qian, Mingang Chen, Lizhuang Ma:
Real-IAD: A Real-World Multi-View Dataset for Benchmarking Versatile Industrial Anomaly Detection. 22883-22892
- Hongwei Zheng, Linyuan Zhou, Han Li, Jinming Su, Xiaoming Wei, Xiaoming Xu:
BEM: Balanced and Entropy-Based Mix for Long-Tailed Semi-Supervised Learning. 22893-22903
- Rudra P. K. Poudel, Harit Pandya, Stephan Liwicki, Roberto Cipolla:
ReCoRe: Regularized Contrastive Representation Learning of World Model. 22904-22913
- Hossein Mirzaei, Mojtaba Nafez, Mohammad Jafari, Mohammad Bagher Soltani, Mohammad Azizmalayeri, Jafar Habibi, Mohammad Sabokrou, Mohammad Hossein Rohban:
Universal Novelty Detection Through Adaptive Contrastive Learning. 22914-22923
- Lukas Knobel, Tengda Han, Yuki M. Asano:
Learning to Count Without Annotations. 22924-22934
- Xiao Zheng, Xiaoshui Huang, Guofeng Mei, Yuenan Hou, Zhaoyang Lyu, Bo Dai, Wanli Ouyang, Yongshun Gong:
Point Cloud Pre-Training with Diffusion Models. 22935-22945
- Ruyi An, Yewen Li, Xu He, Pengjie Gu, Mengchen Zhao, Dong Li, Jianye Hao, Chaojie Wang, Bo An, Mingyuan Zhou:
Improving Unsupervised Hierarchical Representation With Reinforcement Learning. 22946-22956
- Jie Xu, Yazhou Ren, Xiaolong Wang, Lei Feng, Zheng Zhang, Gang Niu, Xiaofeng Zhu:
Investigating and Mitigating the Side Effects of Noisy Views for Self-Supervised Clustering Algorithms in Practical Multi-View Scenarios. 22957-22966
- Zhaowen Li, Yousong Zhu, Zhiyang Chen, Zongxin Gao, Rui Zhao, Chaoyang Zhao, Ming Tang, Jinqiao Wang:
Self-Supervised Representation Learning from Arbitrary Scenarios. 22967-22977
- Chunghyun Park, Seungwook Kim, Jaesik Park, Minsu Cho:
Learning SO(3)-Invariant Semantic Correspondence via Local Shape Transform. 22978-22987
- Yipeng Gao, Zeyu Wang, Wei-Shi Zheng, Cihang Xie, Yuyin Zhou:
Sculpting Holistic 3D Representation in Contrastive Language-Image-3D Pre-Training. 22998-23008
- Jinyang Liu, Wondmgezahu Teshome, Sandesh Ghimire, Mario Sznaier, Octavia I. Camps:
Solving Masked Jigsaw Puzzles with Diffusion Vision Transformers. 23009-23018
- Hao Yan, Zhihui Ke, Xiaobo Zhou, Tie Qiu, Xidong Shi, Dadong Jiang:
DS-NeRV: Implicit Neural Video Representation with Decomposed Static and Dynamic Codes. 23019-23029
- Huzheng Yang, James Gee, Jianbo Shi:
Brain Decodes Deep Nets. 23030-23040
- Siddharth Tourani, Ahmed Alwheibi, Arif Mahmood, Muhammad Haris Khan:
Pose-Guided Self-Training with Two-Stage Clustering for Unsupervised Landmark Discovery. 23041-23051
- Yanhao Wu, Tong Zhang, Wei Ke, Congpei Qiu, Sabine Süsstrunk, Mathieu Salzmann:
Mitigating Object Dependencies: Improving Point Cloud Self-Supervised Learning Through Object Exchange. 23052-23061
- Ke Fan, Zechen Bai, Tianjun Xiao, Tong He, Max Horn, Yanwei Fu, Francesco Locatello, Zheng Zhang:
Adaptive Slot Attention: Object Discovery with Dynamic Slot Number. 23062-23071



- Ruixuan Xiao, Lei Feng, Kai Tang, Junbo Zhao, Yixuan Li, Gang Chen, Haobo Wang:
Targeted Representation Alignment for Open-World Semi-Supervised Learning. 23072-23082
- Morteza Haghir Chehreghani, Mostafa Haghir Chehreghani:
Hierarchical Correlation Clustering and Tree Preserving Embedding. 23083-23093
- Sua Choi, Dahyun Kang, Minsu Cho:
Contrastive Mean-Shift Learning for Generalized Category Discovery. 23094-23104
- Shahaf Arica, Or Rubin, Sapir Gershov, Shlomi Laufer:
CuVLER: Enhanced Unsupervised Object Discoveries through Exhaustive Self-Supervised Transformers. 23105-23114
- Drew A. Hudson, Daniel Zoran, Mateusz Malinowski, Andrew K. Lampinen, Andrew Jaegle, James L. McClelland, Loic Matthey, Felix Hill, Alexander Lerchner:
SODA: Bottleneck Diffusion Models for Representation Learning. 23115-23127
- Linglin Jing, Yiming Ding, Yunpeng Gao, Zhigang Wang, Xu Yan, Dong Wang, Gerald Schaefer, Hui Fang, Bin Zhao, Xuelong Li:
HPL-ESS: Hybrid Pseudo-Labeling for Unsupervised Event-based Semantic Segmentation. 23128-23137
- Lin Long, Haobo Wang, Zhijie Jiang, Lei Feng, Chang Yao, Gang Chen, Junbo Zhao:
Positive-Unlabeled Learning by Latent Group-Aware Meta Disambiguation. 23138-23147
- Jing Ma, Xiang Xiang, Ke Wang, Yuchuan Wu, Yongbin Li:
Aligning Logits Generatively for Principled Black-Box Knowledge Distillation. 23148-23157
- Jiahong Wang, Yinwei Du, Stelian Coros, Bernhard Thomaszewski:
Neural Modes: Self-supervised Learning of Nonlinear Modal Subspaces. 23158-23167
- Yingqi Liu, Yifan Shi, Baoyuan Wu, Qinglun Li, Xueqian Wang, Li Shen:
Decentralized Directed Collaboration for Personalized Federated Learning. 23168-23178
- Jiaming Zhuo, Feiyang Qin, Can Cui, Kun Fu, Bingxin Niu, Mengzhu Wang, Yuanfang Guo, Chuan Wang, Zhen Wang, Xiaochun Cao, Liang Yang:
Improving Graph Contrastive Learning via Adaptive Positive Sampling. 23179-23187
- Tung Le, Khai Nguyen, Shanlin Sun, Nhat Ho, Xiaohui Xie:
Integrating Efficient Optimal Transport and Functional Maps for Unsupervised Shape Correspondence Learning. 23188-23198
- Yunhui Guo, Youren Zhang, Yubei Chen, Stella X. Yu:
Unsupervised Feature Learning with Emergent Data-Driven Prototypicality. 23199-23208
- Vladan Stojnic, Yannis Kalantidis, Giorgos Tolias:
Label Propagation for Zero-shot Classification with Vision-Language Models. 23209-23218
- Jiazu Yu, Yunzhi Zhuge, Lu Zhang, Ping Hu, Dong Wang, Huchuan Lu, You He:
Boosting Continual Learning of Vision-Language Models via Mixture-of-Experts Adapters. 23219-23230
- Yanshuo Wang, Ali Cheraghian, Zeeshan Hayder, Jie Hong, Sameera Ramasinghe, Shafin Rahman, David Ahmedt-Aristizabal, Xuesong Li, Lars Petersson, Mehrtash Harandi:
Backpropagation-free Network for 3D Test-time Adaptation. 23231-23241
- Yun-Yun Tsai, Fu-Chen Chen, Albert Y. C. Chen, Junfeng Yang, Che-Chun Su, Min Sun, Cheng-Hao Kuo:
GDA: Generalized Diffusion for Robust Test-Time Adaptation. 23242-23251
- Yuwen Tan, Qinhao Zhou, Xiang Xiang, Ke Wang, Yuchuan Wu, Yongbin Li:
Semantically-Shifted Incremental Adapter-Tuning is A Continual ViTransformer. 23252-23262
- Zhongqi Yue, Pan Zhou, Richang Hong, Hanwang Zhang, Qianru Sun:
Few-Shot Learner Parameterization by Diffusion Time-Steps. 23263-23272
- Yongxian Wei, Zixuan Hu, Zhenyi Wang, Li Shen, Chun Yuan, Dacheng Tao:
Free: Faster and Better Data-Free Meta-Learning. 23273-23282
- Jiequan Cui, Beier Zhu, Xin Wen, Xiaojuan Qi, Bei Yu, Hanwang Zhang:
Classes Are Not Equal: An Empirical Study on Image Recognition Fairness. 23283-23292
- Jer Pelhan, Alan Lukezic, Vitjan Zavrtanik, Matej Kristan:
DAVE - A Detect-and-Verify Paradigm for Low-Shot Counting. 23293-23302
- Zhimin Yuan, Wankang Zeng, Yanfei Su, Weiquan Liu, Ming Cheng, Yulan Guo, Cheng Wang:
Density-guided Translator Boosts Synthetic-to-Real Unsupervised Domain Adaptive Segmentation of 3D Point Clouds. 23303-23312
- Dinh Phat Do, Taehoon Kim, Jaemin Na, Jiwon Kim, Keonho Lee, Kyunghwan Cho , Wonjun Hwang:
D3T: Distinctive Dual-Domain Teacher Zigzagging Across RGB-Thermal Gap for Domain-Adaptive Object Detection. 23313-23322
- Yuwei Tang, Zhenyi Lin, Qilong Wang, Pengfei Zhu, Qinghua Hu:
AMU-Tuning: Effective Logit Bias for CLIP-based Few-shot Learning. 23323-23333
- Sanqing Qu, Tianpei Zou, Lianghua He, Florian Röhrbein, Alois Knoll, Guang Chen, Changjun Jiang:
LEAD: Learning Decomposition for Source-free Universal Domain Adaptation. 23334-23343



- Yapeng Li, Yong Luo, Zengmao Wang, Bo Du:
Improving Generalized Zero-Shot Learning by Exploring the Diverse Semantics from External Class Names. 23344-23353
- Jayeon Yoo, Dongkwan Lee, Inseop Chung, Donghyun Kim, Nojun Kwak:
What, How, and When Should Object Detectors Update in Continually Changing Test Domains? 23354-23363
- Xinyao Li, Yuke Li, Zhekai Du, Fengling Li, Ke Lu, Jingjing Li:
Split to Merge: Unifying Separated Modalities for Unsupervised Domain Adaptation. 23364-23374
- Zhekai Du, Xinyao Li, Fengling Li, Ke Lu, Lei Zhu, Jingjing Li:
Domain-Agnostic Mutual Prompting for Unsupervised Domain Adaptation. 23375-23384
- Haojie Zhang, Yongyi Su, Xun Xu, Kui Jia:
Improving the Generalization of Segmentation Foundation Model under Distribution Shift via Weakly Supervised Adaptation. 23385-23395
- Harsh Rangwani, Pradipto Mondal, Pradipto Mondal, Mayank Mishra, Ashish Ramayee Asokan, R. Venkatesh Babu:
DeiT-LT: Distillation Strikes Back for Vision Transformer Training on Long-Tailed Datasets. 23396-23406
- Senqiao Yang, Zhuotao Tian, Li Jiang, Jiaya Jia:
Unified Language-Driven Zero-Shot Domain Adaptation. 23407-23415
- Dong Zhao, Shuang Wang, Qi Zang, Licheng Jiao, Nicu Sebe, Zhun Zhong:
Stable Neighbor Denoising for Source-free Domain Adaptive Segmentation. 23416-23427
- Mohammad Fahes, Tuan-Hung Vu, Andrei Bursuc, Patrick Pérez, Raoul de Charette:
A Simple Recipe for Language-Guided Domain Generalized Segmentation. 23428-23437
- Hantao Yao, Rui Zhang, Changsheng Xu:
TCP: Textual-Based Class-Aware Prompt Tuning for Visual-Language Model. 23438-23448
- Jan-Martin O. Steitz, Stefan Roth:
Adapters Strike Back. 23449-23459
- Maorong Wang, Nicolas Michel, Ling Xiao, Toshihiko Yamasaki:
Improving Plasticity in Online Continual Learning via Collaborative Learning. 23460-23469
- Mir Rayat Imtiaz Hossain, Mennatullah Siam, Leonid Sigal, James J. Little:
Visual Prompting for Generalized Few-shot Segmentation: A Multi-scale Approach. 23470-23480
- Shin'ya Yamaguchi, Sekitoshi Kanai, Kazuki Adachi , Daiki Chijiwa:
Adaptive Random Feature Regularization on Fine-tuning Deep Neural Networks. 23481-23490
- Khoi Duc Nguyen, Chen Li, Gim Hee Lee:
ESCAPE: Encoding Super-keypoints for Category-Agnostic Pose Estimation. 23491-23500
- Zining Chen, Weiqiu Wang, Zhicheng Zhao, Fei Su, Aidong Men, Hongying Meng:
PracticalDG: Perturbation Distillation on Vision-Language Models for Hybrid Domain Generalization. 23501-23511
- Zhaorui Tan, Xi Yang, Kaizhu Huang:
Rethinking Multi-Domain Generalization with A General Learning Objective. 23512-23522
- Yuyin Zhou, Xianhang Li, Fengze Liu, Qingyue Wei, Xuxi Chen, Lequan Yu, Cihang Xie, Matthew P. Lungren, Lei Xing:
L2B: Learning to Bootstrap Robust Models for Combating Label Noise. 23523-23533
- Junjie Chen, Jiebin Yan, Yuming Fang, Li Niu:
Meta-Point Learning and Refining for Category-Agnostic Pose Estimation. 23534-23543
- Geunhyeok Yu, Hyoseok Hwang:
A2XP: Towards Private Domain Generalization. 23544-23553
- Da-Wei Zhou, Hai-Long Sun, Han-Jia Ye, De-Chuan Zhan:
Expandable Subspace Ensemble for Pre-Trained Model-Based Class-Incremental Learning. 23554-23564
- Yanpeng Sun, Jiahui Chen, Shan Zhang, Xinyu Zhang, Qiang Chen, Gang Zhang, Errui Ding, Jingdong Wang, Zechao Li:
VRP-SAM: SAM with Visual Reference Prompt. 23565-23574
- Yixiong Zou, Yicong Liu, Yiman Hu, Yuhua Li, Ruixuan Li:
Flatten Long-Range Loss Landscapes for Cross-Domain Few-Shot Learning. 23575-23584
- Boyang Peng, Sanqing Qu, Yong Wu, Tianpei Zou, Lianghua He, Alois Knoll, Guang Chen, Changjun Jiang:
MAP: MAsk-Pruning for Source-Free Model Intellectual Property Protection. 23585-23594
- De Cheng, Zhipeng Xu, Xinyang Jiang, Nannan Wang, Dongsheng Li, Xinbo Gao:
Disentangled Prompt Representation for Domain Generalization. 23595-23604
- Jonas Herzog:
Adapt Before Comparison: A New Perspective on Cross-Domain Few-Shot Segmentation. 23605-23615
- Anurag Roy, Riddhiman Moulick, Vinay Kumar Verma, Saptarshi Ghosh, Abir Das:
Convolutional Prompting meets Language Models for Continual Learning. 23616-23626
- Wenjin Hou, Shiming Chen, Shuhuang Chen, Ziming Hong, Yan Wang, Xuetao Feng, Salman H. Khan, Fahad Shahbaz Khan, Xinge You:
Visual-Augmented Dynamic Semantic Prototype for Generative Zero-Shot Learning. 23627-23637



- Yan-Shuo Liang, Wu-Jun Li:
InfLoRA: Interference-Free Low-Rank Adaptation for Continual Learning. 23638-23647
- Haifeng Xia, Siyu Xia, Zhengming Ding:
Discriminative Pattern Calibration Mechanism for Source-Free Domain Adaptation. 23648-23658
- Mustafa Burak Gurbuz, Jean Michael Moorman, Constantine Dovrolis:
NICE: Neurogenesis Inspired Contextual Encoding for Replay-free Class Incremental Learning. 23659-23669
- Hongwei Yan, Liyuan Wang, Kaisheng Ma, Yi Zhong:
Orchestrate Latent Expertise: Advancing Online Continual Learning with Multi-Level Supervision and Reverse Self-Distillation. 23670-23680
- Julio Silva-Rodríguez, Sina Hajimiri, Ismail Ben Ayed, Jose Dolz:
A Closer Look at the Few-Shot Adaptation of Large Vision-Language Models. 23681-23690
- Chamuditha Jayanga Galappaththige, Sanoojan Baliah, Malitha Gunawardhana, Muhammad Haris Khan:
Towards Generalizing to Unseen Domains with Few Labels. 23691-23700
- Jing Ma:
Improved Self-Training for Test-Time Adaptation. 23701-23710
- Song Tang, Wenxin Su, Mao Ye, Xiatian Zhu:
Source-Free Domain Adaptation with Frozen Multimodal Foundation Model. 23711-23720
- Haipeng Xiong, Angela Yao:
Deep Imbalanced Regression via Hierarchical Classification Adjustment. 23721-23730
- Xu Yang, Xuan Chen, Moqi Li, Kun Wei, Cheng Deng:
A Versatile Framework for Continual Test-Time Domain Adaptation: Balancing Discriminability and Generalizability. 23731-23740
- Yuhang He, Yingjie Chen, Yuhan Jin, Songlin Dong, Xing Wei, Yihong Gong:
DYSON: Dynamic Feature Space Self-Organization for Online Task-Free Class Incremental Learning. 23741-23751
- Ke Fan, Tong Liu, Xingyu Qiu, Yikai Wang, Lian Huai, Zeyu Shangguan, Shuang Gou, Fengjian Liu, Yuqian Fu, Yanwei Fu, Xingqun Jiang:
Test-Time Linear Out-of-Distribution Detection. 23752-23761
- Weizhao He, Yang Zhang, Wei Zhuo, Linlin Shen, Jiaqi Yang, Songhe Deng, Liang Sun:
APSeg: Auto-Prompt Network for Cross-Domain Few-Shot Semantic Segmentation. 23762-23772
- Yunshi Huang, Fereshteh Shakeri, Jose Dolz, Malik Boudiaf, Houda Bahig, Ismail Ben Ayed:
LP++: A Surprisingly Strong Linear Probe for Few-Shot CLIP. 23773-23782
- Maxime Zanella, Ismail Ben Ayed:
On the Test-Time Zero-Shot Generalization of Vision-Language Models: Do we Really need Prompt Learning? 23783-23793
- Rashindrie Perera, Saman K. Halgamuge:
Discriminative Sample-Guided and Parameter-Efficient Feature Space Adaptation for Cross-Domain Few-Shot Learning. 23794-23804
- Pehuen Moure, Longbiao Cheng, Joachim Ott, Zuowen Wang, Shih-Chii Liu:
Regularized Parameter Uncertainty for Improving Generalization in Reinforcement Learning. 23805-23814
- George Eskandar:
An Empirical Study of the Generalization Ability of Lidar 3D Object Detectors to Unseen Domains. 23815-23825
- Lingxiao Yang, Ru-Yuan Zhang, Yanchen Wang, Xiaohua Xie:
MMA: Multi-Modal Adapter for Vision-Language Models. 23826-23837
- Chulin Xie, De-An Huang, Wenda Chu, Daguang Xu, Chaowei Xiao, Bo Li, Anima Anandkumar:
Perada: Parameter-Efficient Federated Learning Personalization with Generalization Guarantees. 23838-23848
- Yibo Miao, Yu Lei, Feng Zhou, Zhijie Deng:
Bayesian Exploration of Pre-Trained Models for Low-Shot Image Classification. 23849-23859
- Minh-Tuan Tran , Trung Le, Xuan-May Thi Le, Mehrtash Harandi, Quan Hung Tran, Dinh Q. Phung:
NAYER: Noisy Layer Data Generation for Efficient and Effective Data-free Knowledge Distillation. 23860-23869
- Minh-Tuan Tran , Trung Le, Xuan-May Le, Mehrtash Harandi, Dinh Phung:
Text-Enhanced Data-Free Approach for Federated Class-Incremental Learning. 23870-23880
- Keon-Hee Park, Kyungwoo Song, Gyeong-Moon Park:
Pre-trained Vision and Language Transformers are Few-Shot Incremental Learners. 23881-23890
- Hyuck Lee, Heeyoung Kim:
CDMAD: Class-Distribution-Mismatch-Aware Debiasing for Class-Imbalanced Semi-Supervised Learning. 23891-23900
- Yige Yuan, Bingbing Xu, Liang Hou, Fei Sun, Huawei Shen, Xueqi Cheng:
TEA: Test-Time Energy Adaptation. 23901-23911
- Wenyu Zhang, Qingmu Liu, Felix Ong Wei Cong, Mohamed Ragab, Chuan-Sheng Foo:
Universal Semi-Supervised Domain Adaptation by Mitigating Common-Class Bias. 23912-23921



- Sravanti Addepalli, Ashish Ramayee Asokan, Lakshay Sharma, R. Venkatesh Babu:
Leveraging Vision-Language Models for Improving Domain Generalization in Image Classification. 23922-23932
- Minhyuk Seo, Hyunseo Koh, Wonje Jeung, Minjae Lee, San Kim, Hankook Lee, Sungjun Cho, Sungik Choi, Hyunwoo Kim, Jonghyun Choi:
Learning Equi-Angular Representations for Online Continual Learning. 23933-23942
- Seun-An Choe, Ah-Hyung Shin, Keon-Hee Park, Jinwoo Choi, Gyeong-Moon Park:
Open-Set Domain Adaptation for Semantic Segmentation. 23943-23953
- Xialei Liu, Jiang-Tian Zhai, Andrew D. Bagdanov, Ke Li, Ming-Ming Cheng:
Task-Adaptive Saliency Guidance for Exemplar-Free Class Incremental Learning. 23954-23963
- Shiming Chen, Wenjin Hou, Salman H. Khan, Fahad Shahbaz Khan:
Progressive Semantic-Guided Vision Transformer for Zero-Shot Learning. 23964-23974
- Zhengqing Gao, Xu-Yao Zhang, Cheng-Lin Liu:
Unified Entropy Optimization for Open-Set Test-Time Adaptation. 23975-23984
- Rishub Tamirisa, Chulin Xie, Wenzuan Bao, Andy Zhou, Ron Arel, Aviv Shamsian:
FedSelect: Personalized Federated Learning with Customized Selection of Parameters for Fine-Tuning. 23985-23994
- Yutian Luo, Shiqi Zhao, Haoran Wu, Zhiwu Lu:
Dual-Enhanced Coreset Selection with Class-Wise Collaboration for Online Blurry Class Incremental Learning. 23995-24004
- Siteng Huang, Biao Gong, Yutong Feng, Min Zhang, Yiliang Lv, Donglin Wang:
Troika: Multi-Path Cross-Modal Traction for Compositional Zero-Shot Learning. 24005-24014
- Fuli Wan, Han Zhao, Xu Yang, Cheng Deng:
Unveiling the Unknown: Unleashing the Power of Unknown to Known in Open-Set Source-Free Domain Adaptation. 24015-24024
- Zihuan Qiu, Yi Xu, Fanman Meng, Hongliang Li, Linfeng Xu, Qingbo Wu:
Dual-Consistency Model Inversion for Non-Exemplar Class Incremental Learning. 24025-24035
- Jiapeng Su, Qi Fan, Wenjie Pei, Guangming Lu, Fanglin Chen:
Domain-Rectifying Adapter for Cross-Domain Few-Shot Segmentation. 24036-24045
- Wenzuan Zhang, Paul Janson, Rahaf Aljundi, Mohamed Elhoseiny:
Overcoming Generic Knowledge Loss with Selective Parameter Update. 24046-24056
- Ali Abbasi, Parsa Nooralinejad, Hamed Pirsavash, Soheil Kolouri:
BrainWash: A Poisoning Attack to Forget in Continual Learning. 24057-24066
- Bolin Ni, Hongbo Zhao, Chenghao Zhang, Ke Hu, Gaofeng Meng, Zhaoxiang Zhang, Shiming Xiang:
Enhancing Visual Continual Learning with Language-Guided Supervision. 24068-24077
- Hao Yang, Liyuan Pan, Yan Yang, Richard I. Hartley, Miaomiao Liu:
LDP: Language-driven Dual-Pixel Image Defocus Deblurring Network. 24078-24087
- Zhen Long, Qiyuan Wang, Yazhou Ren, Yipeng Liu, Ce Zhu:
S²MVTC: A Simple Yet Efficient Scalable Multi-View Tensor Clustering. 24088-24097
- Eric Marcus, Ray Sheombarsing, Jan-Jakob Sonke, Jonas Teuwen:
Task-Driven Wavelets Using Constrained Empirical Risk Minimization. 24098-24107
- Yuchuan Tian, Hanting Chen, Chao Xu, Yunhe Wang:
Image Processing GNN: Breaking Rigidity in Super-Resolution. 24108-24117
- Tianshu Huang, John Miller, Akarsh Prabhakara, Tao Jin, Tarana Laroia, Zico Kolter, Anthony Rowe:
DART: Implicit Doppler Tomography for Radar Novel View Synthesis. 24118-24129
- Prafull Sharma, Varun Jampani, Yuanzhen Li, Xuhui Jia, Dmitry Lagun, Frédo Durand, Bill Freeman, Mark J. Matthews:
Alchemist: Parametric Control of Material Properties with Diffusion Models. 24130-24141
- Zhengqi Li, Richard Tucker, Noah Snavely, Aleksander Holynski:
Generative Image Dynamics. 24142-24153
- Daniel Geng, Inbum Park, Andrew Owens:
Visual Anagrams: Generating Multi-View Optical Illusions with Diffusion Models. 24154-24163
- Keyu Wu, Lingchen Yang, Zhiyi Kuang, Yao Feng, Xutao Han, Yuefan Shen, Hongbo Fu, Kun Zhou, Youyi Zheng:
MonoHair: High-Fidelity Hair Modeling from a Monocular Video. 24164-24173
- Tero Karras, Miika Aittala, Jaakko Lehtinen, Janne Hellsten, Timo Aila, Samuli Laine:
Analyzing and Improving the Training Dynamics of Diffusion Models. 24174-24184
- Zhe Chen, Jiannan Wu, Wenhui Wang, Weijie Su, Guo Chen, Sen Xing, Muyan Zhong, Qinglong Zhang, Xizhou Zhu, Lewei Lu, Bin Li, Ping Luo, Tong Lu, Yu Qiao, Jifeng Dai:
Intern VL: Scaling up Vision Foundation Models and Aligning for Generic Visual-Linguistic Tasks. 24185-24198
- Lisa Dunlap, Yuhui Zhang, Xiaohan Wang, Ruiqi Zhong, Trevor Darrell, Jacob Steinhardt, Joseph E. Gonzalez, Serena Yeung-Levy:
Describing Differences in Image Sets with Natural Language. 24199-24208



- Yusuf Dalva, Pinar Yanardag:
NoiseCLR: A Contrastive Learning Approach for Unsupervised Discovery of Interpretable Directions in Diffusion Models. 24209-24218
- Yixin Liu, Chenrui Fan, Yutong Dai, Xun Chen, Pan Zhou, Lichao Sun:
MetaCloak: Preventing Unauthorized Subject-Driven Text-to-Image Diffusion-Based Synthesis via Meta-Learning. 24219-24228
- Jinbae Im, JeongYeon Nam, Nokyung Park, Hyungmin Lee, Seunghyun Park:
EGTR: Extracting Graph from Transformer for Scene Graph Generation. 24229-24238
- Jiawang Bai, Kuofeng Gao, Shaobo Min, Shu-Tao Xia, Zhifeng Li, Wei Liu:
BadCLIP: Trigger-Aware Prompt Learning for Backdoor Attacks on CLIP. 24239-24250
- Hassan Mahmood, Ehsan Elhamifar:
Semantic-Aware Multi-Label Adversarial Attacks. 24251-24262
- Yuhang Zhou, Zhongyun Hua:
Defense without Forgetting: Continual Adversarial Defense with Anisotropic & Isotropic Pseudo Replay. 24263-24272
- Rongyi Zhu, Zeliang Zhang, Zhuo Liu, Chenliang Xu, Susan Liang:
Learning to Transform Dynamically for Better Adversarial Transferability. 24273-24283
- Xiaopei Zhu, Yuqiu Liu, Zhanhao Hu, Jianmin Li, Xiaolin Hu:
Infrared Adversarial Car Stickers. 24284-24293
- Jiahao Lu, Xingyi Yang, Xinchao Wang:
Unsegment Anything by Simulating Deformation. 24294-24304
- Dong-Dong Wu, Chilin Fu, Weichang Wu, Wenwen Xia, Xiaolu Zhang, Jun Zhou, Min-Ling Zhang:
Efficient Model Stealing Defense with Noise Transition Matrix. 24305-24315
- Yunlong Zhao, Xiaoheng Deng, Yijing Liu, Xinjun Pei, Jiazhi Xia, Wei Chen:
Fully Exploiting Every Real Sample: SuperPixel Sample Gradient Model Stealing. 24316-24325
- Tianrui Lou, Xiaojun Jia, Jindong Gu, Li Liu, Siyuan Liang, Bangyan He, Xiaochun Cao:
Hide in Thicket: Generating Imperceptible and Rational Adversarial Perturbations on 3D Point Clouds. 24326-24335
- Kunyu Wang, Xuanran He, Wenxuan Wang, Xiaosen Wang:
Boosting Adversarial Transferability by Block Shuffle and Rotation. 24336-24346
- Linyu Tang, Lei Zhang:
Robust Overfitting Does Matter: Test-Time Adversarial Purification with FGSM. 24347-24356
- Jinghuai Zhang, Hongbin Liu, Jinyuan Jia, Neil Zhenqiang Gong:
Data Poisoning Based Backdoor Attacks to Contrastive Learning. 24357-24366
- Siyang Wu, Jiakai Wang, Jiejie Zhao, Yazhe Wang, Xianglong Liu:
NAPGuard: Towards Detecting Naturalistic Adversarial Patches. 24367-24376
- Bowen Tang, Zheng Wang, Yi Bin, Qi Dou, Yang Yang, Heng Tao Shen:
Ensemble Diversity Facilitates Adversarial Transferability. 24377-24386
- K. Naveen Kumar, Reshma Mitra, C. Krishna Mohan:
Revamping Federated Learning Security from a Defender's Perspective: A Unified Defense with Homomorphic Encrypted Data Space. 24387-24397
- Zhengyue Zhao, Jinhao Duan, Kaidi Xu, Chenan Wang, Rui Zhang, Zidong Du, Qi Guo, Xing Hu:
Can Protective Perturbation Safeguard Personal Data from Being Exploited by Stable Diffusion? 24398-24407
- Lin Li, Haoyan Guan, Jianing Qiu, Michael W. Spratling:
One Prompt Word is Enough to Boost Adversarial Robustness for Pre-Trained Vision-Language Models. 24408-24419
- Peifei Zhu, Tsubasa Takahashi, Hirokatsu Kataoka:
Watermark-embedded Adversarial Examples for Copyright Protection against Diffusion Models. 24420-24430
- Sheng Yang, Jiawang Bai, Kuofeng Gao, Yong Yang, Yiming Li, Shu-Tao Xia:
Not All Prompts Are Secure: A Switchable Backdoor Attack Against Pre-trained Vision Transformers. 24431-24441
- Qian Li, Yuxiao Hu, Yinpeng Dong, Dongxiao Zhang, Yuntian Chen:
Focus on Hiders: Exploring Hidden Threats for Enhancing Adversarial Training. 24442-24451
- Junhao Zheng, Chenhao Lin, Jiahao Sun, Zhengyu Zhao, Qian Li, Chao Shen:
Physical 3D Adversarial Attacks against Monocular Depth Estimation in Autonomous Driving. 24452-24461
- Ling Lo, Cheng Yu Yeo, Hong-Han Shuai, Wen-Huang Cheng:
Distraction is All You Need: Memory-Efficient Image Immunization against Diffusion-Based Image Editing. 24462-24471
- Lihua Jing, Rui Wang, Wenqi Ren, Xin Dong, Cong Zou:
PAD: Patch-Agnostic Defense against Adversarial Patch Attacks. 24472-24481
- Jaewon Jung, Hongsun Jang, Jaeyong Song, Jinho Lee:
PeerAiD: Improving Adversarial Distillation from a Specialized Peer Tutor. 24482-24491



- Xinli Yue, Ningping Mou, Qian Wang, Lingchen Zhao:
Revisiting Adversarial Training Under Long-Tailed Distributions. 24492-24501
- Sibo Wang, Jie Zhang, Zheng Yuan, Shiguang Shan:
Pre-Trained Model Guided Fine-Tuning for Zero-Shot Adversarial Robustness. 24502-24511
- Yao Huang, Yinpeng Dong, Shouwei Ruan, Xiao Yang, Hang Su, Xingxing Wei:
Towards Transferable Targeted 3D Adversarial Attack in the Physical World. 24512-24522
- Boheng Li, Yishuo Cai, Haowei Li, Feng Xue, Zhifeng Li, Yiming Li:
Nearest is Not Dearest: Towards Practical Defense Against Quantization-Conditioned Backdoor Attacks. 24523-24533
- Jingyao Xu, Yuetong Lu, Yandong Li, Siyang Lu, Dongdong Wang, Xiang Wei:
Perturbing Attention Gives You More Bang for the Buck: Subtle Imaging Perturbations That Efficiently Fool Customized Diffusion Models. 24534-24543
- Xiangyu Yin, Wenjie Ruan:
Boosting Adversarial Training via Fisher-Rao Norm-Based Regularization. 24544-24553
- Huihui Gong, Minjing Dong, Siqi Ma, Seyit Camtepe, Surya Nepal, Chang Xu:
Random Entangled Tokens for Adversarially Robust Vision Transformer. 24554-24563
- Jiyang Guan, Jian Liang, Ran He:
Backdoor Defense via Test-Time Detecting and Repairing. 24564-24573
- Bernd Prach, Fabio Brau, Giorgio C. Buttazzo, Christoph H. Lampert:
1-Lipschitz Layers Compared: Memory, Speed, and Certifiable Robustness. 24574-24583
- Yuhao Sun, Lingyun Yu, Hongtao Xie, Jiaming Li, Yongdong Zhang:
DiffAM: Diffusion-Based Adversarial Makeup Transfer for Facial Privacy Protection. 24584-24594
- Amira Guesmi, Ruitian Ding, Muhammad Abdullah Hanif, Ihsen Alouani, Muhammad Shafique:
DAP: A Dynamic Adversarial Patch for Evading Person Detectors. 24595-24604
- Shenglin Yin, Zhen Xiao, Mingxuan Song, Jieyi Long:
Adversarial Distillation Based on Slack Matching and Attribution Region Alignment. 24605-24614
- Han Wu, Guanyan Ou, Weibin Wu, Zibin Zheng:
Improving Transferable Targeted Adversarial Attacks with Model Self-Enhancement. 24615-24624
- Xuanming Cui, Alejandro Aparcedo, Young Kyun Jang, Ser-Nam Lim:
On the Robustness of Large Multimodal Models Against Image Adversarial Attacks. 24625-24634
- Takami Sato, Justin Yue, Nanze Chen, Ningfei Wang, Qi Alfred Chen:
Intriguing Properties of Diffusion Models: An Empirical Study of the Natural Attack Capability in Text-to-Image Generative Models. 24635-24644
- Siyuan Liang, Mingli Zhu, Aishan Liu, Baoyuan Wu, Xiaochun Cao, Ee-Chien Chang:
BadCLIP: Dual-Embedding Guided Backdoor Attack on Multimodal Contrastive Learning. 24645-24654
- Yanting Wang, Hongye Fu, Wei Zou, Jinyuan Jia:
MMCert: Provable Defense Against Adversarial Attacks to Multi-Modal Models. 24655-24664
- Kaiyu Song, Hanjiang Lai, Yan Pan, Jian Yin:
MimicDiffusion: Purifying Adversarial Perturbation via Mimicking Clean Diffusion Model. 24665-24674
- Zeyu Wang, Xianhang Li, Hongru Zhu, Cihang Xie:
Revisiting Adversarial Training at Scale. 24675-24685
- Xiao Li, Wei Zhang, Yining Liu, Zhanhao Hu, Bo Zhang, Xiaolin Hu:
Language-Driven Anchors for Zero-Shot Adversarial Robustness. 24686-24695
- Di Ming, Peng Ren, Yunlong Wang, Xin Feng:
Transferable Structural Sparse Adversarial Attack Via Exact Group Sparsity Training. 24696-24705
- Zhuoxiao Li, Zhihang Zhong, Shohei Nobuhara, Ko Nishino, Yinqiang Zheng:
Fooling Polarization-Based Vision Using Locally Controllable Polarizing Projection. 24706-24715
- Erh-Chung Chen, Pin-Yu Chen, I-Hsin Chung, Che-Rung Lee:
Overload: Latency Attacks on Object Detection for Edge Devices. 24716-24725
- Samar Fares, Karthik Nandakumar:
Attack To Defend: Exploiting Adversarial Attacks for Detecting Poisoned Models. 24726-24735
- Samyak Jain, Tania Dutta:
Towards Understanding and Improving Adversarial Robustness of Vision Transformers. 24736-24745
- Yanghao Zhang, Tianle Zhang, Ronghui Mu, Xiaowei Huang, Wenjie Ruan:
Towards Fairness-Aware Adversarial Learning. 24746-24755
- Peng Sun, Xinyang Liu, Zhibo Wang, Bo Liu:
Byzantine-robust Decentralized Federated Learning via Dual-domain Clustering and Trust Bootstrapping. 24756-24765
- Yuan Xiao, Shiqing Ma, Juan Zhai, Chunrong Fang, Jinyuan Jia, Zhenyu Chen:
Towards General Robustness Verification of MaxPool-Based Convolutional Neural Networks via Tightening Linear Approximation. 24766-24775



- Zhuorong Li, Daiwei Yu, Lina Wei, Canghong Jin, Yun Zhang, Sixian Chan:
Soften to Defend: Towards Adversarial Robustness via Self-Guided Label Refinement. 24776-24785
- K. L. Navaneet, Soroush Abbasi Koohpayegani, Essam Sleiman, Hamed Pirsiavash:
SlowFormer: Adversarial Attack on Compute and Energy Consumption of Efficient Vision Transformers. 24786-24797
- Siyuan Cheng, Guanhong Tao, Yingqi Liu, Guangyu Shen, Shengwei An, Shiwei Feng, Xiangzhe Xu, Kaiyuan Zhang, Shiqing Ma, Xiangyu Zhang:
Lotus: Evasive and Resilient Backdoor Attacks through Sub-Partitioning. 24798-24809
- Sabbir Ahmed, Ranyang Zhou, Shaahin Angizi, Adnan Siraj Rakin:
Deep-TROJ: An Inference Stage Trojan Insertion Algorithm Through Efficient Weight Replacement Attack. 24810-24819
- Alvi Md. Ishmam, Christopher Thomas:
Semantic Shield: Defending Vision-Language Models Against Backdooring and Poisoning via Fine-Grained Knowledge Alignment. 24820-24830
- Andong Hua, Jindong Gu, Zhiyu Xue, Nicholas Carlini, Eric Wong, Yao Qin:
Initialization Matters for Adversarial Transfer Learning. 24831-24840
- Zhengwei Fang, Rui Wang, Tao Huang, Liping Jing:
Strong Transferable Adversarial Attacks via Ensembled Asymptotically Normal Distribution Learning. 24841-24850
- Gangwei Xu, Yujin Wang, Jinwei Gu, Tianfan Xue, Xin Yang:
HDRFlow: Real-Time HDR Video Reconstruction with Large Motions. 24851-24860
- Jin Gong, Runzhao Yang, Weihang Zhang, Jinli Suo, Qionghai Dai:
A Physics-Informed Low-Rank Deep Neural Network for Blind and Universal Lens Aberration Correction. 24861-24870
- Yanchen Dong, Ruiqin Xiong, Jian Zhang, Zhaofei Yu, Xiaopeng Fan, Shuyuan Zhu, Tiejun Huang:
Super-Resolution Reconstruction from Bayer-Pattern Spike Streams. 24871-24880
- Xin Wang, Lizhi Wang, Xiangtian Ma, Maoqing Zhang, Lin Zhu, Hua Huang:
In2SET: Intra-Inter Similarity Exploiting Transformer for Dual-Camera Compressive Hyperspectral Imaging. 24881-24891
- Teng Hu, Ran Yi, Baihong Qian, Jiangning Zhang, Paul L. Rosin, Yu-Kun Lai:
SuperSVG: Superpixel-Based Scalable Vector Graphics Synthesis. 24892-24901
- Hao Yang, Liyuan Pan, Yan Yang, Wei Liang:
Language-driven All-in-one Adverse Weather Removal. 24902-24912
- Haofeng Zhong, Yuchen Hong, Shuchen Weng, Jinxiu Liang, Boxin Shi:
Language-guided Image Reflection Separation. 24913-24922
- Shuji Habuchi, Keita Takahashi, Chihiro Tsutake, Toshiaki Fujii, Hajime Nagahara:
Time-Efficient Light-Field Acquisition Using Coded Aperture and Events. 24923-24933
- Yifei Xia, Chu Zhou, Chengxuan Zhu, Minggui Teng, Chao Xu, Boxin Shi:
NB-GTR: Narrow-Band Guided Turbulence Removal. 24934-24943
- Jianping Jiang, Xinyu Zhou, Bingxuan Wang, Xiaoming Deng, Chao Xu, Boxin Shi:
Complementing Event Streams and RGB Frames for Hand Mesh Reconstruction. 24944-24954
- Rui Zhao, Ruiqin Xiong, Jing Zhao, Jian Zhang, Xiaopeng Fan, Zhaofei Yu, Tiejun Huang:
Boosting Spike Camera Image Reconstruction from a Perspective of Dealing with Spike Fluctuations. 24955-24965
- Taewoo Kim, Hoonhee Cho, Kuk-Jin Yoon:
Frequency-Aware Event-Based Video Deblurring for Real-World Motion Blur. 24966-24976
- Yixin Yang, Jinxiu Liang, Bohan Yu, Yan Chen, Jimmy S. Ren, Boxin Shi:
Latency Correction for Event-Guided Deblurring and Frame Interpolation. 24977-24986
- Jiaqi Tang, Ruizheng Wu, Xiaogang Xu, Sixing Hu, Ying-Cong Chen:
Learning to Remove Wrinkled Transparent Film with Polarized Prior. 24987-24996
- Suhyun Shin, Seokjun Choi, Felix Heide, Seung-Hwan Baek:
Dispersed Structured Light for Hyperspectral 3D Imaging. 24997-25006
- Varun Sundar, Matthew Dutson, Andrei Ardelean, Claudio Bruschini, Edoardo Charbon, Mohit Gupta:
Generalized Event Cameras. 25007-25017
- Changqing Su, Zhiyuan Ye, Yongsheng Xiao, You Zhou, Zhen Cheng, Bo Xiong, Zhaofei Yu, Tiejun Huang:
Intensity-Robust Autofocus for Spike Camera. 25018-25027
- Krzysztof A. Maliszewski, Magdalena A. Urbanska, Varvara Vetrova, Sylwia M. Kolenderska:
Selective Nonlinearities Removal from Digital Signals. 25028-25036
- Seunghyun Shin, Jihwan Bae, Jisu Shin, Inwook Shim, Hae-Gon Jeon:
Close Imitation of Expert Retouching for Black-and-White Photography. 25037-25046
- Jiyuan Zhang, Shiyan Chen, Yajing Zheng, Zhaofei Yu, Tiejun Huang:
Spike-guided Motion Deblurring with Unknown Modal Spatiotemporal Alignment. 25047-25057



- Wei-Yu Chen, Aswin C. Sankaranarayanan, Anat Levin, Matthew O'Toole:
Coherence as Texture - Passive Textureless 3D Reconstruction by Self-Interference. 25058-25066
- Parsa Mirdehghan, Maxx Wu, Wenzheng Chen, David B. Lindell, Kiriakos N. Kutulakos:
TurboSL: Dense, Accurate and Fast 3D by Neural Inverse Structured Light. 25067-25076
- Tomoki Ichikawa, Shohei Nobuhara, Ko Nishino:
SPIDeRS: Structured Polarization for Invisible Depth and Reflectance Sensing. 25077-25085
- Zhen Guo, Hongping Gan:
CPP-Net: Embracing Multi-Scale Feature Fusion into Deep Unfolding CP-PPA Network for Compressive Sensing. 25086-25095
- Hoon Kim, Minje Jang, Wonjun Yoon, Jisoo Lee, Donghyun Na, Sanghyun Woo:
SwitchLight: Co-Design of Physics-Driven Architecture and Pre-training Framework for Human Portrait Relighting. 25096-25106
- Dong Lao, Congli Wang, Alex Wong, Stefano Soatto:
Diffeomorphic Template Registration for Atmospheric Turbulence Mitigation. 25107-25116
- Yakun Chang, Yeliduosi Xiaokaiti, Yujia Liu , Bin Fan, Zhaojun Huang, Tiejun Huang, Boxin Shi:
Towards HDR and HFR Video from Rolling-Mixed-Bit Spikings. 25117-25127
- Chong Wang, Lanqing Guo, Yufei Wang, Hao Cheng, Yi Yu, Bihan Wen:
Progressive Divide-and-Conquer via Subsampling Decomposition for Accelerated MRI. 25128-25137
- Vishal Purohit, Junjie Luo, Yiheng Chi, Qi Guo, Stanley H. Chan, Qiang Qiu:
Generative Quanta Color Imaging. 25138-25148
- Xiaoyang Wang, Hongping Gan:
UFC-Net: Unrolling Fixed-point Continuous Network for Deep Compressive Sensing. 25149-25159
- Zhicheng Cai, Hao Zhu, Qiu Shen, Xinran Wang, Xun Cao:
Batch Normalization Alleviates the Spectral Bias in Coordinate Networks. 25160-25171
- Rui Jiang, Fangwen Tu, Yixuan Long, Aabhaas Vaish, Bowen Zhou, Qinyi Wang, Wei Zhang, Yuntan Fang, Luis Eduardo Garcia Capel, Bo Mu, Tiejun Dai, Andreas Suess:
EVS-Assisted Joint Deblurring, Rolling-Shutter Correction and Video Frame Interpolation Through Sensor Inverse Modeling. 25172-25181
- Zhile Chen, Yuhui Quan, Hui Ji:
Unsupervised Deep Unrolling Networks for Phase Unwrapping. 25182-25192
- Changjin Kim, Tae Hyun Kim, Sungyong Baik:
LAN: Learning to Adapt Noise for Image Denoising. 25193-25202
- Sarah Friday, Yunzi Shi , Yaswanth Cherivirala, Vishwanath Saragadam, Adithya Pedireddla:
Snapshot Lidar: Fourier Embedding of Amplitude and Phase for Single-Image Depth Reconstruction. 25203-25212
- Haobo Xu, Jun Zhou, Hua Yang, Renjie Pan, Cunyan Li:
FC-GNN: Recovering Reliable and Accurate Correspondences from Interferences. 25213-25222
- Mark Sheinin, Aswin C. Sankaranarayanan, Srinivasa G. Narasimhan:
Projecting Trackable Thermal Patterns for Dynamic Computer Vision. 25223-25232
- Haley M. So, Laurie Bose, Piotr Dudek, Gordon Wetzstein:
PixelRNN: In-pixel Recurrent Neural Networks for End-to-end-optimized Perception with Neural Sensors. 25233-25244
- Tomer Garber, Tom Tirer:
Image Restoration by Denoising Diffusion Models with Iteratively Preconditioned Guidance. 25245-25254
- Matthieu Terris, Thomas Moreau, Nelly Pustelnik, Julián Tachella:
Equivariant Plug-and-Play Image Reconstruction. 25255-25264
- Sachin Shah, Matthew A. Chan, Haoming Cai, Jingxi Chen, Sakshum Kulshrestha, Chahat Deep Singh, Yiannis Aloimonos, Christopher A. Metzler:
CodedEvents: Optimal Point-Spread-Function Engineering for 3D-Tracking with Event Cameras. 25265-25275
- Mingyang Xie, Haiyun Guo, Brandon Y. Feng, Lingbo Jin, Ashok Veeraraghavan, Christopher A. Metzler:
WaveMo: Learning Wavefront Modulations to See Through Scattering. 25276-25285
- Ripon Kumar Saha, Dehao Qin, Nianyi Li, Jinwei Ye, Suren Jayasuriya:
Turb-Seg-Res: A Segment-then-Restore Pipeline for Dynamic Videos with Atmospheric Turbulence. 25286-25296
- Zhenghao Pan, Haijin Zeng, Jiezhang Cao, Kai Zhang, Yongyong Chen:
DiffSCI: Zero-Shot Snapshot Compressive Imaging via Iterative Spectral Diffusion Model. 25297-25306
- Stanley H. Chan, Hashan K. Weerasooriya, Weijian Zhang, Pamela Abshire, István Gyöngy, Robert K. Henderson:
Resolution Limit of Single-Photon LiDAR. 25307-25316
- Ishak Ayad, Nicolas Larue, Maï K. Nguyen:
QN-Mixer: A Quasi-Newton MLP-Mixer Model for Sparse-View CT Reconstruction. 25317-25326
- Gang Qu, Ping Wang, Xin Yuan:
Dual-Scale Transformer for Large-Scale Single-Pixel Imaging. 25327-25337



- Mingdeng Cao, Sidi Yang, Yujiu Yang, Yinqiang Zheng:
Rolling Shutter Correction with Intermediate Distortion Flow Estimation. 25338-25347
- Bhargav Ghanekar, Salman Siddique Khan, Pranav Sharma, Shreyas Singh, Vivek Boominathan, Kaushik Mitra, Ashok Veeraraghavan:
Passive Snapshot Coded Aperture Dual-Pixel RGB-D Imaging. 25348-25357
- Brandon Zhao, Aviad Levis, Liam Connor, Pratul P. Srinivasan, Katherine L. Bouman:
Single View Refractive Index Tomography with Neural Fields. 25358-25367
- Zhiyang Yao, Shuyang Liu, Xiaoyun Yuan, Lu Fang:
SPECAT: SPatial-spEctral Cumulative-Attention Transformer for High-Resolution Hyperspectral Image Reconstruction. 25368-25377
- Xiaoqian Lv, Shengping Zhang, Chenyang Wang, Yichen Zheng, Bineng Zhong, Chongyi Li, Liqiang Nie:
Fourier Priors-Guided Diffusion for Zero-Shot Joint Low-Light Enhancement and Deblurring. 25378-25388
- Yiyu Li, Ke Xu, Gerhard Petrus Hancke, Rynson W. H. Lau:
Color Shift Estimation-and-Correction for Image Enhancement. 25389-25398
- Xingyu Zhou, Leheng Zhang, Xiaorui Zhao, Keze Wang, Leida Li, Shuhang Gu:
Video Super-Resolution Transformer with Masked Inter&Intra-Frame Attention. 25399-25408
- Quan Zhang, Xiaoyu Liu, Wei Li, Hanting Chen, Junchao Liu, Jie Hu, Zhiwei Xiong, Chun Yuan, Yunhe Wang:
Distilling Semantic Priors from SAM to Efficient Image Restoration Models. 25409-25419
- Xianyu Chen, Ming Jiang, Qi Zhao:
Beyond Average: Individualized Visual Scanpath Prediction. 25420-25431
- Yang Ai, Huabo Huang, Xiaoqiang Zhou, Jiexiang Wang, Ran He:
Multimodal Prompt Perceiver: Empower Adaptiveness, Generalizability and Fidelity for All-in-One Image Restoration. 25432-25444
- Dian Zheng, Xiao-Ming Wu, Shuzhou Yang, Jian Zhang, Jian-Fang Hu, Wei-Shi Zheng:
Selective Hourglass Mapping for Universal Image Restoration Based on Diffusion Model. 25445-25455
- Rongyuan Wu, Tao Yang, Lingchen Sun, Zhengqiang Zhang, Shuai Li, Lei Zhang:
SeeSR: Towards Semantics-Aware Real-World Image Super-Resolution. 25456-25467
- Yurui Zhu, Xueyang Fu, Peng-Tao Jiang, Hao Zhang, Qibin Sun, Jinwei Chen, Zheng-Jun Zha, Bo Li:
Revisiting Single Image Reflection Removal in the Wild. 25468-25478
- Zhongze Wang, Haitao Zhao, Jingchao Peng, Lujian Yao, Kaijie Zhao:
ODCR: Orthogonal Decoupling Contrastive Regularization for Unpaired Image Dehazing. 25479-25489
- Haoning Wu, Zicheng Zhang, Erli Zhang, Chaofeng Chen, Liang Liao, Annan Wang, Kaixin Xu, Chunyi Li, Jingwen Hou, Guangtao Zhai, Geng Xue, Wenxiu Sun, Qiong Yan, Weisi Lin:
Q-Instruct: Improving Low-Level Visual Abilities for Multi-Modality Foundation Models. 25490-25500
- Qunliang Xing, Mai Xu, Shengxi Li, Xin Deng, Meisong Zheng, Huaida Liu, Ying Chen:
Enhancing Quality of Compressed Images by Mitigating Enhancement Bias Towards Compression Domain. 25501-25511
- Dongyoung Kim, Jinwoo Kim, Junsang Yu, Seon Joo Kim:
Attentive Illumination Decomposition Model for Multi-Illuminant White Balancing. 25512-25521
- Shuwei Li, Robby T. Tan:
NightCC: Nighttime Color Constancy via Adaptive Channel Masking. 25522-25531
- Hongjun Wang, Jiyuan Chen, Yinqiang Zheng, Tieyong Zeng:
Navigating Beyond Dropout: An Intriguing Solution Towards Generalizable Image Super Resolution. 25532-25543
- Yuekun Dai, Shangchen Zhou, Qinyue Li, Chongyi Li, Chen Change Loy:
Learning Inclusion Matching for Animation Paint Bucket Colorization. 25544-25553
- Yujia Liu , Chenxi Yang, Dingquan Li, Jianhao Ding, Tingting Jiang:
Defense Against Adversarial Attacks on No-Reference Image Quality Models with Gradient Norm Regularization. 25554-25563
- Zhihao Duan, Ming Lu, Justin Yang, Jiangpeng He, Zhan Ma, Fengqing Zhu:
Towards Backward-Compatible Continual Learning of Image Compression. 25564-25573
- Boyang Wang, Fengyu Yang, Xihang Yu, Chao Zhang, Hanbin Zhao:
APISR: Anime Production Inspired Real-World Anime Super-Resolution. 25574-25584
- Zixuan Ye, Wenze Liu, He Guo, Yujia Liang, Chaoyi Hong, Hao Lu, Zhiguo Cao:
Unifying Automatic and Interactive Matting with Pretrained ViTs. 25585-25594
- Chengxu Liu, Xuan Wang, Xiangyu Xu, Ruhao Tian, Shuai Li, Xueming Qian, Ming-Hsuan Yang:
Motion-Adaptive Separable Collaborative Filters for Blind Motion Deblurring. 25595-25605
- Yijun Yang, Hongtao Wu, Angelica I. Avilés-Rivero, Yulun Zhang, Jing Qin, Lei Zhu:
Genuine Knowledge from Practice: Diffusion Test-Time Adaptation for Video Adverse Weather Removal. 25606-25616
- Jie Xiao, Xueyang Fu, Yurui Zhu, Dong Li, Jie Huang, Kai Zhu, Zheng-Jun Zha:
HomoFormer: Homogenized Transformer for Image Shadow Removal. 25617-25626



- Xiang Chen, Jinshan Pan, Jiangxin Dong:
Bidirectional Multi-Scale Implicit Neural Representations for Image Deraining. 25627-25636
- Yuxing Duan:
LED: A Large-scale Real-world Paired Dataset for Event Camera Denoising. 25637-25647
- Haoyue Liu, Shihan Peng, Lin Zhu, Yi Chang, Hanyu Zhou, Luxin Yan:
Seeing Motion at Nighttime with an Event Camera. 25648-25658
- Chen Zhang, Wencheng Han, Yang Zhou, Jianbing Shen, Cheng-Zhong Xu, Wentao Liu:
Leveraging Frame Affinity for sRGB-to-RawVideo De-Rendering. 25659-25668
- Fanghua Yu, Jinjin Gu, Zheyuan Li, Jinfan Hu, Xiangtao Kong, Xintao Wang, Jingwen He, Yu Qiao, Chao Dong:
Scaling Up to Excellence: Practicing Model Scaling for Photo-Realistic Image Restoration In the Wild. 25669-25680
- Xintian Mao, Qingli Li, Yan Wang:
AdaRevD: Adaptive Patch Exiting Reversible Decoder Pushes the Limit of Image Deblurring. 25681-25690
- Lufei Chen, Xiangpeng Tian, Shuhua Xiong, Yinjie Lei, Chao Ren:
Unsupervised Blind Image Deblurring Based on Self-Enhancement. 25691-25700
- Hoonhee Cho, Taewoo Kim, Yuhwan Jeong, Kuk-Jin Yoon:
TTA-EVF: Test-Time Adaptation for Event-based Video Frame Interpolation via Reliable Pixel and Sample Estimation. 25701-25711
- Longguang Wang, Juncheng Li, Yingqian Wang, Qingyong Hu, Yulan Guo:
Learning Coupled Dictionaries from Unpaired Data for Image Super-Resolution. 25712-25721
- Wei Yu, Jie Huang, Bing Li, Kaiwen Zheng, Qi Zhu, Man Zhou, Feng Zhao:
Empowering Resampling Operation for Ultra-High-Definition Image Enhancement with Model-Aware Guidance. 25722-25731
- Tao Hu, Qingsen Yan, Yuankai Qi, Yanning Zhang:
Generating Content for HDR Deghosting from Frequency View. 25732-25741
- Jiancheng Zhang, Haijin Zeng, Jiezhang Cao, Yongyong Chen, Dengxiu Yu, Yin-Ping Zhao:
Dual Prior Unfolding for Snapshot Compressive Imaging. 25742-25752
- Gengchen Zhang, Yulun Zhang, Xin Yuan, Ying Fu:
Binarized Low-Light Raw Video Enhancement. 25753-25762
- Ilya Chugunov, David Shustein, Ruyu Yan, Chenyang Lei, Felix Heide:
Neural Spline Fields for Burst Image Fusion and Layer Separation. 25763-25773
- Yanhui Guo, Fangzhou Luo, Xiaolin Wu:
Learning Degradation-Independent Representations for Camera ISP Pipelines. 25774-25783
- Bingchen Li, Xin Li, Hanxin Zhu, Yeying Jin, Ruoyu Feng, Zhizheng Zhang, Zhibo Chen:
SeD: Semantic-Aware Discriminator for Image Super-Resolution. 25784-25795
- Yufei Wang, Wenhan Yang, Xinyuan Chen, Yaohui Wang, Lanqing Guo, Lap-Pui Chau, Ziwei Liu, Yu Qiao, Alex C. Kot, Bihan Wen:
SinSR: Diffusion-Based Image Super-Resolution in a Single Step. 25796-25805
- Qingping Zheng, Ling Zheng, Yuanfan Guo, Ying Li, Songcen Xu, Jiankang Deng, Hang Xu:
Self-Adaptive Reality-Guided Diffusion for Artifact-Free Super-Resolution. 25806-25816
- Jiancheng Zhang, Haijin Zeng, Yongyong Chen, Dengxiu Yu, Yin-Ping Zhao:
Improving Spectral Snapshot Reconstruction with Spectral-Spatial Rectification. 25817-25826
- Yuzhe Zhang, Jiawei Zhang, Hao Li, Zhouxia Wang, Luwei Hou, Dongqing Zou, Liheng Bian:
Diffusion-based Blind Text Image Super-Resolution. 25827-25836
- Yan Wang, Yi Liu, Shijie Zhao, Junlin Li, Li Zhang:
CAMixerSR: Only Details Need More "Attention". 25837-25846
- Jia-Hao Wu, Fu-Jen Tsai, Yan-Tsung Peng, Chung-Chi Tsai, Chia-Wen Lin, Yen-Yu Lin:
ID-Blau: Image Deblurring by Implicit Diffusion-Based reBLurring AUGmentation. 25847-25856
- Haoyu Chen, Wenbo Li, Jinjin Gu, Jingjing Ren, Haoze Sun, Xueyi Zou, Zhensong Zhang, Youliang Yan, Lei Zhu:
Low-Res Leads the Way: Improving Generalization for Super-Resolution by Self-Supervised Learning. 25857-25867
- Haoze Sun, Wenbo Li, Jianzhuang Liu, Haoyu Chen, Renjing Pei, Xueyi Zou, Youliang Yan, Yujiu Yang:
CoSeR: Bridging Image and Language for Cognitive Super-Resolution. 25868-25878
- Insoo Kim, Jaeseok Choi, Geonseok Seo, Kinam Kwon, Jinwoo Shin, Hyong-Euk Lee:
Real-World Efficient Blind Motion Deblurring via Blur Pixel Discretization. 25879-25888
- Dihan Zheng, Yihang Zou, Xiaowen Zhang, Chenglong Bao:
SeNM-VAE: Semi-Supervised Noise Modeling with Hierarchical Variational Autoencoder. 25889-25899
- Kanchana Vaishnavi Gandikota, Paramanand Chandramouli:
Text-Guided Explorable Image Super-Resolution. 25900-25911
- Zixiang Zhao, Haowen Bai, Jiangshe Zhang, Yulun Zhang, Kai Zhang, Shuang Xu, Dongdong Chen, Radu Timofte, Luc Van Gool:
Equivariant Multi-Modality Image Fusion. 25912-25921



- Jiangtong Tan, Jie Huang, Naishan Zheng, Man Zhou, Keyu Yan, Danfeng Hong, Feng Zhao:
Revisiting Spatial-Frequency Information Integration from a Hierarchical Perspective for Panchromatic and Multi-Spectral Image Fusion. 25922-25931
- Haokai Zhu, Si-Yuan Cao, Jianxin Hu, Sitong Zuo, Beinan Yu, Jiacheng Ying, Junwei Li, Hui-Liang Shen:
MCNet: Rethinking the Core Ingredients for Accurate and Efficient Homography Estimation. 25932-25941
- Ziyu Shan, Yujie Zhang, Qi Yang, Haichen Yang, Yiling Xu, Jenq-Neng Hwang, Xiaozhong Xu, Shan Liu:
Contrastive Pre-Training with Multi-View Fusion for No-Reference Point Cloud Quality Assessment. 25942-25951
- Caixia Zhou, Yaping Huang, Mengyang Pu, Qingji Guan, Ruoxi Deng, Haibin Ling:
MuGE: Multiple Granularity Edge Detection. 25952-25962
- Yiting Lu, Xin Li, Yajing Pei, Kun Yuan, Qizhi Xie, Yunpeng Qu, Ming Sun, Chao Zhou, Zhibo Chen:
KVQ: Kwai Video Quality Assessment for Short-form Videos. 25963-25973
- Jun Cheng , Dong Liang, Shan Tan:
Transfer CLIP for Generalizable Image Denoising. 25974-25984
- Kexuan Shi, Xingyu Zhou, Shuhang Gu:
Improved Implicit Neural Representation with Fourier Reparameterized Training. 25985-25994
- Yuyao Ye, Ning Zhang, Yang Zhao, Hongbin Cao, Ronggang Wang:
Deep Video Inverse Tone Mapping Based on Temporal Clues. 25995-26004
- Li-Yuan Tsao, Yi-Chen Lo, Chia-Che Chang, Hao-Wei Chen, Roy Tseng, Chien Feng, Chun-Yi Lee:
Boosting Flow-based Generative Super-Resolution Models via Learned Prior. 26005-26015
- Yinglong Li, Jiacheng Li, Zhiwei Xiong:
Look-Up Table Compression for Efficient Image Restoration. 26016-26025
- Zongyao He, Zhi Jin:
Latent Modulated Function for Computational Optimal Continuous Image Representation. 26026-26035
- Xingtong Ge, Jixiang Luo, Xinjie Zhang, Tongda Xu, Guo Lu, Dailan He, Jing Geng, Yan Wang, Jun Zhang, Hongwei Qin:
Task-Aware Encoder Control for Deep Video Compression. 26036-26045
- Zhixiong Yang, Jingyuan Xia, Shengxi Li, Xinghua Huang, Shuanghui Zhang, Zhen Liu, Yaowen Fu, Yongxiang Liu:
A Dynamic Kernel Prior Model for Unsupervised Blind Image Super-Resolution. 26046-26056
- Wenjing Wang, Huan Yang, Jianlong Fu, Jiaying Liu:
Zero-Reference Low-Light Enhancement via Physical Quadruple Priors. 26057-26066
- Woohyeok Kim, Geonu Kim, Junyong Lee, Seungyong Lee, Seung-Hwan Baek, Sunghyun Cho:
ParamISP: Learned Forward and Inverse ISPs Using Camera Parameters. 26067-26076
- Xianzu Wu, Xianfeng Wu, Tianyu Luan, Yajing Bai, Zhongyuan Lai, Junsong Yuan:
FSC: Few-Point Shape Completion. 26077-26087
- Zhaoyang Jia, Jiahao Li, Bin Li, Houqiang Li, Yan Lu:
Generative Latent Coding for Ultra-Low Bitrate Image Compression. 26088-26098
- Jiahao Li, Bin Li, Yan Lu:
Neural Video Compression with Feature Modulation. 26099-26108
- Junkai Fan, Jiangwei Weng, Kun Wang, Yijun Yang, Jianjun Qian, Jun Li, Jian Yang:
Driving-Video Dehazing with Non-Aligned Regularization for Safety Assistance. 26109-26119
- Abhisek Ray, Gaurav Kumar, Maheshkumar H. Kolekar:
CFAT: Unleashing Triangular Windows for Image Super-resolution. 26120-26129
- Ruoxi Zhu, Shusong Xu, Peiye Liu, Sicheng Li, Yanheng Lu, Dimin Niu, Zihao Liu, Zihao Meng, Zhiyong Li, Xinhua Chen, Yibo Fan:
Zero-Shot Structure-Preserving Diffusion Model for High Dynamic Range Tone Mapping. 26130-26139
- Chenyu You, Yifei Min, Weicheng Dai, Jasjeet S. Sekhon, Lawrence H. Staib, James S. Duncan:
Calibrating Multi-modal Representations: A Pursuit of Group Robustness without Annotations. 26140-26150
- Suyuan Liu, Ke Liang, Zhibin Dong, Siwei Wang, Xihong Yang, Sihang Zhou, En Zhu, Xinwang Liu:
Learn from View Correlation: An Anchor Enhancement Strategy for Multi-View Clustering. 26151-26161
- Hao Xiong, Yehui Tang, Xinyu Ye, Junchi Yan:
Circuit Design and Efficient Simulation of Quantum Inner Product and Empirical Studies of Its Effect on Near-Term Hybrid Quantum-Classic Machine Learning. 26162-26170
- Yue Yuan, Rundong He, Yicong Dong, Zhongyi Han, Yilong Yin:
Discriminability-Driven Channel Selection for Out-of-Distribution Detection. 26171-26180
- Jiantong Jiang, Zeyi Wen, Atif Bin Mansoor, Ajmal Mian:
Efficient Hyperparameter Optimization with Adaptive Fidelity Identification. 26181-26190
- Jan-Nico Zaech, Martin Danelljan, Tolga Birdal, Luc Van Gool:
Probabilistic Sampling of Balanced K-Means using Adiabatic Quantum Computing. 26191-26201
- Fei Ye, Adrian G. Bors:
Online Task-Free Continual Generative and Discriminative Learning via Dynamic Cluster Memory. 26202-26212



- Xin Zhang, Jiawei Du, Yunsong Li, Weiying Xie, Joey Tianyi Zhou:
Spanning Training Progress: Temporal Dual-Depth Scoring (TDDS) for Enhanced Dataset Pruning. 26213-26222
- Yuan Wang, Huazhu Fu, Renuga Kanagavelu, Qingsong Wei, Yong Liu, Rick Siow Mong Goh:
An Aggregation-Free Federated Learning for Tackling Data Heterogeneity. 26223-26232
- Jiayi Guan, Li Shen, Ao Zhou, Lusong Li, Han Hu, Xiaodong He, Guang Chen, Changjun Jiang:
POCE: Primal Policy Optimization with Conservative Estimation for Multi-constraint Offline Reinforcement Learning. 26233-26243
- Chong Peng, Pengfei Zhang, Yongyong Chen, Zhao Kang, Chenglizhao Chen, Qiang Shawn Cheng:
Fine-Grained Bipartite Concept Factorization for Clustering. 26254-26264
- Yijun Yang, Tianyi Zhou, Kanxue Li, Dapeng Tao, Lusong Li, Li Shen, Xiaodong He, Jing Jiang, Yuhui Shi:
Embodied Multi-Modal Agent trained by an LLM from a Parallel TextWorld. 26265-26275
- Myeongseob Ko, Feiyang Kang, Weiyan Shi, Ming Jin, Zhou Yu, Ruoxi Jia:
The Mirrored Influence Hypothesis: Efficient Data Influence Estimation by Harnessing Forward Passes. 26276-26285
- Haotian Liu, Chunyuan Li, Yuheng Li, Yong Jae Lee:
Improved Baselines with Visual Instruction Tuning. 26286-26296
- Zheren Fu, Lei Zhang, Hou Xia, Zhendong Mao:
Linguistic-Aware Patch Slimming Framework for Fine-Grained Cross-Modal Alignment. 26297-26306
- Shuai Tan, Bin Ji, Ye Pan:
FlowVQTalker: High-Quality Emotional Talking Face Generation through Normalizing Flow and Quantization. 26307-26317
- Jinxiang Liu, Yikun Liu, Fei Zhang, Chen Ju, Ya Zhang, Yanfeng Wang:
Audio-Visual Segmentation via Unlabeled Frame Exploitation. 26318-26329
- Fengyu Yang, Chao Feng, Ziyang Chen, Hyoungseob Park, Daniel Wang, Yiming Dou, Ziyao Zeng, Xien Chen, Rit Gangopadhyay, Andrew Owens, Alex Wong:
Binding Touch to Everything: Learning Unified Multimodal Tactile Representations. 26330-26343
- Jiawei Ma, Po-Yao Huang, Saining Xie, Shang-Wen Li, Luke Zettlemoyer, Shih-Fu Chang, Wen-Tau Yih, Hu Xu:
MoDE: CLIP Data Experts via Clustering. 26344-26353
- Anna Kukleva, Fadime Sener, Edoardo Remelli, Bugra Tekin, Eric Sauser, Bernt Schiele, Shugao Ma:
X-MIC: Cross-Modal Instance Conditioning for Egocentric Action Generalization. 26354-26363
- Zhongwei Ren, Zhicheng Huang, Yunchao Wei, Yao Zhao, Dongmei Fu, Jiashi Feng, Xiaojie Jin:
PixelLM: Pixel Reasoning with Large Multimodal Model. 26364-26373
- Naishan Zheng, Man Zhou, Jie Huang, Junming Hou, Haoying Li, Yuan Xu, Feng Zhao:
Probing Synergistic High-Order Interaction in Infrared and Visible Image Fusion. 26374-26385
- Wenqi Jia, Miao Liu, Hao Jiang, Ishwarya Ananthabhotla, James M. Rehg, Vamsi Krishna Ithapu, Ruohan Gao:
The Audio-Visual Conversational Graph: From an Egocentric-Exocentric Perspective. 26386-26395
- Yining Hong, Zishuo Zheng, Peihao Chen, Yian Wang, Junyan Li, Chuang Gan:
MultiPLY: A Multisensory Object-Centric Embodied Large Language Model in 3D World. 26396-26406
- Zhangyang Qi, Ye Fang, Zeyi Sun, Xiaoyang Wu, Tong Wu, Jiaqi Wang, Dahua Lin, Hengshuang Zhao:
GPT4Point: A Unified Framework for Point-Language Understanding and Generation. 26407-26417
- Sijin Chen, Xin Chen, Chi Zhang, Mingsheng Li, Gang Yu, Hao Fei, Hongyuan Zhu, Jiayuan Fan, Tao Chen:
LL3DA: Visual Interactive Instruction Tuning for Omni-3D Understanding, Reasoning, and Planning. 26418-26428
- Jiasen Lu, Christopher Clark, Sangho Lee, Zichen Zhang, Savya Khosla, Ryan Marten, Derek Hoiem, Aniruddha Kembhavi:
Unified-IO 2: Scaling Autoregressive Multimodal Models with Vision, Language, Audio, and Action. 26429-26445
- Minghao Chen, Junyu Xie, Iro Laina, Andrea Vedaldi:
Shap-Editor: Instruction-guided Latent 3D Editing in Seconds. 26446-26456
- Dongjin Kim, Sung Jin Um, Sangmin Lee, Jung Uk Kim:
Learning to Visually Localize Sound Sources from Mixtures without Prior Source Knowledge. 26457-26466
- Hanyu Zhou, Yi Chang, Zhiwei Shi:
Bring Event into RGB and LiDAR: Hierarchical Visual-Motion Fusion for Scene Flow. 26467-26476
- Hao Zhang, Linfeng Tang, Xinyu Xiang, Xuhui Zuo, Jiayi Ma:
Dispel Darkness for Better Fusion: A Controllable Visual Enhancer Based on Cross-Modal Conditional Adversarial Learning. 26477-26486
- Yuanhong Chen, Yuyuan Liu, Hu Wang, Fengbei Liu, Chong Wang, Helen Frazer, Gustavo Carneiro:
Unraveling Instance Associations: A Closer Look for Audio-Visual Segmentation. 26487-26497
- Haoran Xu, Peixi Peng, Guang Tan, Yuan Li, Xinhai Xu, Yonghong Tian:
DMR: Decomposed Multi-Modality Representations for Frames and Events Fusion in Visual Reinforcement Learning. 26498-26508
- Mingyu Lee, Jongwon Choi:
Text-Guided Variational Image Generation for Industrial Anomaly Detection and Segmentation. 26509-26518



- Yiming Dou, Fengyu Yang, Yi Liu, Antonio Loquercio, Andrew Owens:
Tactile-Augmented Radiance Fields. 26519-26529
- Gongwei Chen, Leyang Shen, Rui Shao, Xiang Deng, Liqiang Nie:
LION : Empowering Multimodal Large Language Model with Dual-Level Visual Knowledge. 26530-26540
- Xiaojun Hou, Jiazheng Xing, Yijie Qian, Yaowei Guo, Shuo Xin, Junhao Chen, Kai Tang, Mengmeng Wang, Zhengkai Jiang, Liang Liu, Yong Liu:
SDSTrack: Self-Distillation Symmetric Adapter Learning for Multi-Modal Visual Object Tracking. 26541-26551
- Yichi Zhang, Yinpeng Dong, Siyuan Zhang, Tianzan Min, Hang Su, Jun Zhu:
Exploring the Transferability of Visual Prompting for Multimodal Large Language Models. 26552-26562
- Yong Xien Chng, Henry Zheng, Yizeng Han, Xuchong Qiu, Gao Huang:
Mask Grounding for Referring Image Segmentation. 26563-26573
- Jiaming Han, Kaixiong Gong, Yiyuan Zhang, Jiaqi Wang, Kaipeng Zhang, Dahua Lin, Yu Qiao, Peng Gao, Xiangyu Yue:
OneLLM: One Framework to Align All Modalities with Language. 26574-26585
- Hongxia Xie, Chu-Jun Peng, Yu-Wen Tseng, Hung-Jen Chen, Chan-Feng Hsu, Hong-Han Shuai, Wen-Huang Cheng:
EmoVIT: Revolutionizing Emotion Insights with Visual Instruction Tuning. 26586-26595
- Xinyu Wang, Bohan Zhuang, Qi Wu:
ModaVerse: Efficiently Transforming Modalities with LLMs. 26596-26606
- Zheng Li, Xiang Li, Xinyi Fu, Xin Zhang, Weiqiang Wang, Shuo Chen, Jian Yang:
PromptKD: Unsupervised Prompt Distillation for Vision-Language Models. 26607-26616
- Wenyi Mo, Tianyu Zhang, Yalong Bai, Bing Su, Ji-Rong Wen, Qing Yang:
Dynamic Prompt Optimizing for Text-to-Image Generation. 26617-26626
- Qinglong Cao, Zhengqin Xu, Yuntian Chen, M. Chao, Xiaokang Yang:
Domain Prompt Learning with Quaternion Networks. 26627-26636
- Weixian Lei, Yixiao Ge, Kun Yi, Jianfeng Zhang, Difei Gao, Dylan Sun, Yuying Ge, Ying Shan, Mike Zheng Shou:
VIT-LENS: Towards Omni-modal Representations. 26637-26647
- Sihan Liu, Yiwei Ma, Xiaoqing Zhang, Haowei Wang, Jiayi Ji, Xiaoshuai Sun, Rongrong Ji:
Rotated Multi-Scale Interaction Network for Referring Remote Sensing Image Segmentation. 26648-26658
- Zhaojian Li, Bin Zhao, Yuan Yuan:
Cyclic Learning for Binaural Audio Generation and Localization. 26659-26668
- Haochen Han, Qinghua Zheng, Guang Dai, Minnan Luo, Jingdong Wang:
Learning to Rematch Mismatched Pairs for Robust Cross-Modal Retrieval. 26669-26678
- Ji Lin, Hongxu Yin, Wei Ping, Pavlo Molchanov, Mohammad Shoeybi, Song Han:
VILA: On Pre-training for Visual Language Models. 26679-26689
- Jack Urbanek, Florian Bordes, Pietro Astolfi, Mary Williamson, Vasu Sharma, Adriana Romero-Soriano:
A Picture is Worth More Than 77 Text Tokens: Evaluating CLIP-Style Models on Dense Captions. 26690-26699
- Li Li, Jiawei Peng, Huiyi Chen, Chongyang Gao, Xu Yang:
How to Configure Good In-Context Sequence for Visual Question Answering. 26700-26710
- Yuxin Guo, Siyang Sun, Shuailei Ma, Kecheng Zheng, Xiaoyi Bao, Shijie Ma, Wei Zou, Yun Zheng:
CrossMAE: Cross-Modality Masked Autoencoders for Region-Aware Audio-Visual Pre-Training. 26711-26721
- Baochen Xiong, Xiaoshan Yang, Yaguang Song, Yaowei Wang, Changsheng Xu:
Modality-Collaborative Test-Time Adaptation for Action Recognition. 26722-26731
- Tanvir Mahmud, Yapeng Tian, Diana Marculescu:
T-VSL: Text-Guided Visual Sound Source Localization in Mixtures. 26732-26741
- Yuanhuiyi Lyu, Xu Zheng, Jiazhou Zhou, Lin Wang:
UniBind: LLM-Augmented Unified and Balanced Representation Space to Bind Them All. 26742-26752
- Zhang Li, Biao Yang, Qiang Liu, Zhiyin Ma, Shuo Zhang, Jingxu Yang, Yabo Sun, Yuliang Liu, Xiang Bai:
Monkey: Image Resolution and Text Label are Important Things for Large Multi-Modal Models. 26753-26763
- Guanzhou Ke, Bo Wang, Xiaoli Wang, Shengfeng He:
Rethinking Multi-View Representation Learning via Distilled Disentangling. 26764-26773
- Taeheon Kim, Sebin Shin, Youngjoon Yu, Hak Gu Kim, Yong Man Ro:
Causal Mode Multiplexer: A Novel Framework for Unbiased Multispectral Pedestrian Detection. 26774-26783
- Ji-Jia Wu, Andy Chia-Hao Chang, Chieh-Yu Chuang, Chun-Pei Chen, Yu-Lun Liu, Min-Hung Chen, Hou-Ning Hu, Yung-Yu Chuang, Yen-Yu Lin:
Image-Text Co-Decomposition for Text-Supervised Semantic Segmentation. 26784-26793
- A. J. Piergiovanni, Isaac Noble, Dahun Kim, Michael S. Ryoo, Victor Gomes, Anelia Angelova:
Mirasol3B: A Multimodal Autoregressive Model for Time-Aligned and Contextual Modalities. 26794-26804
- Zihao Wei, Zixuan Pan, Andrew Owens:
Efficient Vision-Language Pre-Training by Cluster Masking. 26805-26815
- Sanjoy Chowdhury, Sayan Nag, K. J. Joseph, Balaji Vasan Srinivasan, Dinesh Manocha:
MELFuSION: Synthesizing Music from Image and Language Cues Using Diffusion Models. 26816-26825



- Chen Chen, Jiahao Qi, Xingyue Liu, Kangcheng Bin, Ruigang Fu, Xikun Hu, Ping Zhong:
Weakly Misalignment-Free Adaptive Feature Alignment for UAVs-Based Multimodal Object Detection. 26826-26835
- Clara Fernandez-Labrador, Mertcan Akçay, Eitan Abecassis, Joan Massich, Christopher Schroers:
DiVAS: Video and Audio Synchronization with Dynamic Frame Rates. 26836-26844
- Tian Liang, Jing Huang, Ming Kong, Luyuan Chen, Qiang Zhu:
Querying as Prompt: Parameter-Efficient Learning for Multimodal Language Model. 26845-26855
- Zifeng Xie, Shengye Yu, Qile He, Mengtian Li:
Sonic VisionLM: Playing Sound with Vision Language Models. 26856-26865
- Zixian Gao, Xun Jiang, Xing Xu, Fumin Shen, Yujie Li, Heng Tao Shen:
Embracing Unimodal Aleatoric Uncertainty for Robust Multimodal Fusion. 26866-26875
- Juntao Zhang, Yuehuai Liu, Yu-Wing Tai, Chi-Keung Tang:
C3Net: Compound Conditioned ControlNet for Multimodal Content Generation. 26876-26885
- Omkar Thawakar, Muzammal Naseer, Rao Muhammad Anwer, Salman H. Khan, Michael Felsberg, Mubarak Shah, Fahad Shahbaz Khan:
Composed Video Retrieval via Enriched Context and Discriminative Embeddings. 26886-26896
- Nikhil Singh, Chih-Wei Wu, Iroro Orife, Mahdi M. Kalayeh:
Looking Similar, Sounding Different: Leveraging Counterfactual Cross-Modal Pairs for Audiovisual Representation Learning. 26897-26908
- Jinwei Han, Zhiwen Lin, Zhongyisun Sun, Yingguo Gao, Ke Yan, Shouhong Ding, Yuan Gao, Gui-Song Xia:
Anchor-based Robust Finetuning of Vision-Language Models. 26909-26918
- Mengyue Geng, Lin Zhu, Lizhi Wang, Wei Zhang, Ruiqin Xiong, Yonghong Tian:
Event-Based Visible and Infrared Fusion via Multi-Task Collaboration. 26919-26929
- Jinyoung Park, Juyeon Ko, Hyunwoo J. Kim:
Prompt Learning via Meta-Regularization. 26930-26940
- Yucheng Suo, Fan Ma, Linchao Zhu, Yi Yang:
Knowledge-Enhanced Dual-Stream Zero-Shot Composed Image Retrieval. 26941-26952
- Kaili Sun, Zhiwen Xie, Mang Ye, Huyin Zhang:
Contextual Augmented Global Contrast for Multimodal Intent Recognition. 26953-26963
- Hao Zhang, Xuhui Zuo, Jie Jiang, Chunchao Guo, Jiayi Ma:
MRFS: Mutually Reinforcing Image Fusion and Segmentation. 26964-26973
- Zhenye Luo, Min Ren, Xuecai Hu, Yongzhen Huang, Li Yao:
POPDG: Popular 3D Dance Generation with PopDanceSet. 26974-26983
- Yuxin Chen, Zongyang Ma, Ziqi Zhang, Zhongang Qi, Chunfeng Yuan, Bing Li, Junfu Pu, Ying Shan, Xiaojuan Qi, Weiming Hu:
How to Make Cross Encoder a Good Teacher for Efficient Image-Text Retrieval? 26984-26993
- Jihwan Bang, Sumyeong Ahn, Jae-Gil Lee:
Active Prompt Learning in Vision Language Models. 26994-27004
- Christopher Liao, Theodoros Tsiligkaridis, Brian Kulis:
Descriptor and Word Soups Q: Overcoming the Parameter Efficiency Accuracy Tradeoff for Out-of-Distribution Few-shot Learning. 27005-27015
- Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, Jiayi Ma:
Text-IF: Leveraging Semantic Text Guidance for Degradation-Aware and Interactive Image Fusion. 27016-27025
- Chaoya Jiang, Haiyang Xu, Mengfan Dong, Jiaxing Chen, Wei Ye, Ming Yan, Qinghao Ye, Ji Zhang, Fei Huang, Shikun Zhang:
Hallucination Augmented Contrastive Learning for Multimodal Large Language Model. 27026-27036
- Lei Zhu, Fangyun Wei, Yanye Lu:
Beyond Text: Frozen Large Language Models in Visual Signal Comprehension. 27037-27047
- Sagnik Majumder, Ziad Al-Halah, Kristen Grauman:
Learning Spatial Features from Audio-Visual Correspondence in Egocentric Videos. 27048-27058
- Yuanhang Zhang, Shuang Yang, Shiguang Shan, Xilin Chen:
ES³: Evolving Self-Supervised Learning of Robust Audio-Visual Speech Representations. 27059-27069
- Xu Peng, Junwei Zhu, Boyuan Jiang, Ying Tai, Donghao Luo, Jiangning Zhang, Wei Lin, Taisong Jin, Chengjie Wang, Rongrong Ji:
PortraitBooth: A Versatile Portrait Model for Fast Identity-Preserved Personalization. 27070-27080
- Le Xue, Ning Yu, Shu Zhang, Artemis Panagopoulou, Junnan Li, Roberto Martín-Martín, Jiajun Wu, Caiming Xiong, Ran Xu, Juan Carlos Niebles, Silvio Savarese:
ULIP-2: Towards Scalable Multimodal Pre-Training for 3D Understanding. 27081-27091
- Trevine Oorloff, Surya Koppisetti, Nicolò Bonettini, Divyraj Solanki, Ben Colman, Yaser Yacoob, Ali Shahriyari, Gaurav Bharaj:
AVFF: Audio-Visual Feature Fusion for Video Deepfake Detection. 27092-27102



- Bo Zou, Chao Yang, Yu Qiao, Chengbin Quan, Youjian Zhao:
Language-aware Visual Semantic Distillation for Video Question Answering. 27103-27113
- Renjie Pi, Lewei Yao, Jiahui Gao, Jipeng Zhang, Tong Zhang:
PerceptionGPT: Effectively Fusing Visual Perception Into LLM. 27114-27123
- Qi Yang, Xing Nie, Tong Li, Pengfei Gao, Ying Guo, Cheng Zhen, Pengfei Yan, Shiming Xiang:
Cooperation Does Matter: Exploring Multi-Order Bilateral Relations for Audio-Visual Segmentation. 27124-27133
- Xiaojie Jin, Bowen Zhang, Weibo Gong, Kai Xu, Xueqing Deng, Peng Wang, Zhao Zhang, Xiaohui Shen, Jiashi Feng:
MV-Adapter: Multimodal Video Transfer Learning for Video Text Retrieval. 27134-27143
- Weijian Ma, Shuaiqi Chen, Yunzhong Lou, Xueyang Li, Xiangdong Zhou:
Draw Step by Step: Reconstructing CAD Construction Sequences from Point Clouds via Multimodal Diffusion. 27144-27153
- Anton Ratnarajah, Sreyan Ghosh, Sonal Kumar, Purva Chiniya, Dinesh Manocha:
AV-RIR: Audio-Visual Room Impulse Response Estimation. 27154-27165
- Yan Tai, Weichen Fan, Zhao Zhang, Ziwei Liu:
Link-Context Learning for Multimodal LLMs. 27166-27175
- Shentong Mo, Pedro Morgado:
Unveiling the Power of Audio-Visual Early Fusion Transformers with Dense Interactions Through Masked Modeling. 27176-27186
- Yang Qin, Yingke Chen, Dezhong Peng, Xi Peng, Joey Tianyi Zhou, Peng Hu:
Noisy-Correspondence Learning for Text-to-Image Person Re-Identification. 27187-27196
- Jiaxuan Chen, Yu Qi, Yueming Wang, Gang Pan:
Mind Artist: Creating Artistic Snapshots with Human Thought. 27197-27207
- Kang Chen, Xiangqian Wu:
VTQA: Visual Text Question Answering via Entity Alignment and Cross-Media Reasoning. 27208-27217
- Prannay Kaul, Zhizhong Li, Hao Yang, Yonatan Dukler, Ashwin Swaminathan, C. J. Taylor, Stefano Soatto:
THRONE: An Object-Based Hallucination Benchmark for the Free-Form Generations of Large Vision-Language Models. 27218-27228
- Noël Vouitsis, Zhaoyan Liu, Satya Krishna Gorti, Valentin Villecroze, Jesse C. Cresswell, Guangwei Yu, Gabriel Loaiza-Ganem, Maksims Volkovs:
Data-Efficient Multimodal Fusion on a Single GPU. 27229-27241
- Changan Chen, Kumar Ashutosh, Rohit Girdhar, David Harwath, Kristen Grauman:
SoundingActions: Learning How Actions Sound from Narrated Egocentric Videos. 27242-27252
- Sameera Ramasinghe, Violetta Shevchenko, Gil Avraham, Thalaiyasingam Ajanthan:
Accept the Modality Gap: An Exploration in the Hyperbolic Space. 27253-27262
- Junwen Xiong, Peng Zhang, Tao You, Chuanyue Li, Wei Huang, Yufei Zha:
DiffSal: Joint Audio and Video Learning for Diffusion Saliency Prediction. 27263-27273
- Sikai Bai, Jie Zhang, Song Guo, Shuaicheng Li, Jingcai Guo, Jun Hou, Tao Han, Xiaocheng Lu:
DiPrompT: Disentangled Prompt Tuning for Multiple Latent Domain Generalization in Federated Learning. 27274-27283
- Karren D. Yang, Anurag Ranjan, Jen-Hao Rick Chang, Raviteja Vemulapalli, Oncel Tuzel:
Probabilistic Speech-Driven 3D Facial Motion Synthesis: New Benchmarks, Methods, and Applications. 27284-27293
- Xinyi Jiang, Guoming Wang, Junhao Guo, Juncheng Li, Wenqiao Zhang, Rongxing Lu, Siliang Tang:
DIEM: Decomposition-Integration Enhancing Multimodal Insights. 27294-27303
- Jaeseok Byun, Dohoon Kim, Taesup Moon:
MAFA: Managing False Negatives for Vision-Language Pre-Training. 27304-27314
- Jeongsoo Choi, Se Jin Park, Minsu Kim, Yong Man Ro:
AV2AV: Direct Audio-Visual Speech to Audio-Visual Speech Translation with Unified Audio-Visual Speech Representation. 27315-27327
- Yake Wei, Ruoxuan Feng, Zihe Wang, Di Hu:
Enhancing Multimodal Cooperation via Sample-Level Modality Valuation. 27328-27337
- Sizhe Li, Yiming Qin, Minghang Zheng, Xin Jin, Yang Liu:
Diff-BGM: A Diffusion Model for Video Background Music Generation. 27338-27347
- Sitong Wu, Haoru Tan, Zhuotao Tian, Yukang Chen, Xiaojuan Qi, Jiaya Jia:
SaCo Loss: Sample-Wise Affinity Consistency for Vision-Language Pre-Training. 27348-27359
- Haokun Lin, Haoli Bai, Zhili Liu, Lu Hou, Muyi Sun, Linqi Song, Ying Wei, Zhenan Sun:
MoPE-CLIP: Structured Pruning for Efficient Vision-Language Models with Module-Wise Pruning Error Metric. 27360-27370
- Zihua Zhao, Mengxi Chen, Tianjie Dai, Jiangchao Yao, Bo Han, Ya Zhang, Yanfeng Wang:
Mitigating Noisy Correspondence by Geometrical Structure Consistency Learning. 27371-27380



- Chao Yi, Lu Ren, De-Chuan Zhan, Han-Jia Ye:
Leveraging Cross-Modal Neighbor Representation for Improved CLIP Classification. 27392-27401
- Siddharth Srivastava, Gaurav Sharma:
OmniVec2 - A Novel Transformer Based Network for Large Scale Multimodal and Multitask Learning. 27402-27414
- Zineng Tang, Ziyi Yang, Mahmoud Khademi, Yang Liu, Chenguang Zhu, Mohit Bansal:
CoDi-2: In-Context, Interleaved, and Interactive Any-to-Any Generation. 27415-27424
- Xiaoqiang Yan, Zhixiang Jin, Fengshou Han, Yangdong Ye:
Differentiable Information Bottleneck for Deterministic Multi-View Clustering. 27425-27434
- Yusheng Dai, Hang Chen, Jun Du, Ruoyu Wang, Shihao Chen, Haotian Wang, Chin-Hui Lee:
A Study of Dropout-Induced Modality Bias on Robustness to Missing Video Frames for Audio-Visual Speech Recognition. 27435-27445
- Xiaohui Zhang, Jaehong Yoon, Mohit Bansal, Huaxiu Yao:
Multimodal Representation Learning by Alternating Unimodal Adaptation. 27446-27456
- Shilong Ou, Zhe Xue, Yawen Li, Meiyu Liang, Yuanqiang Cai, Junjiang Wu:
View-Category Interactive Sharing Transformer for Incomplete Multi-View Multi-Label Learning. 27457-27466
- Tianyu Huang, Liangzu Peng, René Vidal, Yun-Hui Liu:
Scalable 3D Registration via Truncated Entry-Wise Absolute Residuals. 27467-27477
- Viktoria Ehm, Maolin Gao, Paul Roetzer, Marvin Eisenberger, Daniel Cremers, Florian Bernard:
Partial-to-Partial Shape Matching with Geometric Consistency. 27478-27487
- Qingyu Song, Wei Lin, Juncheng Wang, Hong Xu:
Towards Robust Learning to Optimize with Theoretical Guarantees. 27488-27496
- Swaminathan Gurumurthy, Karnik Ram, Bingqing Chen, Zachary Manchester, Zico Kolter:
From Variance to Veracity: Unbundling and Mitigating Gradient Variance in Differentiable Bundle Adjustment Layers. 27497-27506
- Nastaran Saadati, Minh Pham, Nasla Saleem, Joshua R. Waite, Aditya Balu, Zhanhong Jiang, Chinmay Hegde, Soumik Sarkar:
DIMAT: Decentralized Iterative Merging-And-Training for Deep Learning Models. 27507-27517
- Hao Jiang, Bingfeng Zhou, Yadong Mu:
Ink Dot-Oriented Differentiable Optimization for Neural Image Halftoning. 27518-27527
- Guobin Shen, Dongcheng Zhao, Tenglong Li, Jindong Li, Yi Zeng:
Are Conventional SNNs Really Efficient? A Perspective from Network Quantization. 27528-27537
- Hong Huang, Weiming Zhuang, Chen Chen, Lingjuan Lyu:
FedMef: Towards Memory-Efficient Federated Dynamic Pruning. 27538-27547
- Xinghui Li, Jingyi Lu, Kai Han, Victor Adrian Prisacariu:
SD4Match: Learning to Prompt Stable Diffusion Model for Semantic Matching. 27548-27558
- Guobiao Li, Sheng Li, Zicong Luo, Zhenxing Qian, Xinpeng Zhang:
Purified and Unified Steganographic Network. 27559-27568
- Zhe Zhang, Huairui Wang, Zhenzhong Chen, Shan Liu:
Learned Lossless Image Compression Based on Bit Plane Slicing. 27569-27578
- Jiacheng Cheng, Nuno Vasconcelos:
Towards Calibrated Multi-Label Deep Neural Networks. 27579-27589
- Nishant Jain, Arun S. Suggala, Pradeep Shenoy:
Improving Generalization via Meta-Learning on Hard Samples. 27590-27599
- Noo-Ri Kim, Jin-Seop Lee, Jee-Hyong Lee:
Learning with Structural Labels for Learning with Noisy Labels. 27600-27610
- Khawar Islam, Muhammad Zaigham Zaheer, Arif Mahmood, Karthik Nandakumar:
Diffusemix: Label-Preserving Data Augmentation with Diffusion Models. 27611-27620
- Yinhua Piao, Sangseon Lee, Yijingxiu Lu, Sun Kim:
Improving Out-of-Distribution Generalization in Graphs via Hierarchical Semantic Environments. 27621-27630
- Shreyas Fadnavis, Agniva Chowdhury, Joshua Batson, Petros Drineas, Eleftherios Garyfallidis:
Patch2Self2: Self-Supervised Denoising on Coresets via Matrix Sketching. 27631-27641
- Junfeng Cheng, Tania Stathaki:
G-FARS: Gradient-Field-Based Auto-Regressive Sampling for 3D Part Grouping. 27642-27651
- Fahimeh Hosseini Noohdani, Parsa Hosseini, Aryan Yazdan Parast, HamidReza Yaghoubi Araghi, Mahdieh Soleymani Baghshah:
Decompose-and-Compose: A Compositional Approach to Mitigating Spurious Correlation. 27652-27661
- Xin Guo, Jiangwei Lao, Bo Dang, Yingying Zhang, Lei Yu, Lixiang Ru, Liheng Zhong, Ziyuan Huang, Kang Wu, Dingxiang Hu, Huimei He, Jian Wang, Jingdong Chen, Ming Yang, Yongjun Zhang, Yansheng Li:
SkySense: A Multi-Modal Remote Sensing Foundation Model Towards Universal Interpretation for Earth Observation Imagery. 27662-27673



- Runmin Dong, Shuai Yuan, Bin Luo, Mengxuan Chen, Jinxiao Zhang, Lixian Zhang, Weijia Li, Juepeng Zheng, Haohuan Fu: **Building Bridges Across Spatial and Temporal Resolutions: Reference-Based Super-Resolution via Change Priors and Conditional Diffusion Model.** 27674-27684
- Aysim Toker, Marvin Eisenberger, Daniel Cremers, Laura Leal-Taixé: **SatSynth: Augmenting Image-Mask Pairs Through Diffusion Models for Aerial Semantic Segmentation.** 27685-27695
- Xuyang Li, Danfeng Hong, Jocelyn Chanussot: **S2MAE: A Spatial-Spectral Pretraining Foundation Model for Spectral Remote Sensing Data.** 27696-27705
- Xinhao Cai, Qiuxia Lai, Yuwei Wang, Wenguan Wang, Zeren Sun, Yazhou Yao: **Poly Kernel Inception Network for Remote Sensing Detection.** 27706-27716
- Zhuohong Li, Wei He, Jiepan Li, Fangxiao Lu, Hongyan Zhang: **Learning without Exact Guidance: Updating Large-Scale High-Resolution Land Cover Maps from Low-Resolution Historical Labels.** 27717-27727
- Weijia Li, Haote Yang, Zhenghao Hu, Juepeng Zheng, Gui-Song Xia, Conghui He: **3D Building Reconstruction from Monocular Remote Sensing Images with Multi-level Supervisions.** 27728-27737
- Yule Duan, Xiao Wu, Haoyu Deng, Liang-Jian Deng: **Content-Adaptive Non-Local Convolution for Remote Sensing Pan sharpening.** 27738-27747
- Junyan Ye, Qiyan Luo, Jinhua Yu, Huaping Zhong, Zhimeng Zheng, Conghui He, Weijia Li: **SG-BEV: Satellite-Guided BEV Fusion for Cross-View Semantic Segmentation.** 27748-27757
- Demin Yu, Xutao Li, Yunming Ye, Baoquan Zhang, Chuyao Luo, Kuai Dai, Rui Wang, Xunlai Chen: **DiffCast: A Unified Framework via Residual Diffusion for Precipitation Nowcasting.** 27758-27767
- Ziyang Chen, Wei Long, He Yao, Yongjun Zhang, Bingshu Wang, Yongbin Qin, Jia Wu: **MoCha-Stereo: Motif Channel Attention Network for Stereo Matching.** 27768-27777
- Shangfeng Huang, Ruisheng Wang, Bo Guo, Hongxin Yang: **PBWR: Parametric-Building-Wireframe Reconstruction from Aerial LiDAR Point Clouds.** 27778-27787
- Vitus Benson, Claire Robin, Christian Requena-Mesa, Lázaro Alonso, Nuno Carvalhais, José Cortés, Zhihan Gao, Nora Linscheid, Mélanie Weynants, Markus Reichstein: **Multi-Modal Learning for Geospatial Vegetation Forecasting.** 27788-27799
- Wenhao Wu, Hau-San Wong, Si Wu, Tianyou Zhang: **Relational Matching for Weakly Semi-Supervised Oriented Object Detection.** 27800-27810
- Mubashir Noman, Muzammal Naseer, Hisham Cholakkal, Rao Muhammad Anwer, Salman H. Khan, Fahad Shahbaz Khan: **Rethinking Transformers Pre-training for Multi-Spectral Satellite Imagery.** 27811-27819
- Haijin Zeng, Jiezhang Cao, Kai Zhang, Yongyong Chen, Hiep Luong, Wilfried Philips: **Unmixing Diffusion for Self-Supervised Hyperspectral Image Denoising.** 27820-27830
- Kartik Kuckreja, Muhammad Sohail Danish, Muzammal Naseer, Abhijit Das, Salman Khan, Fahad Shahbaz Khan: **GeoChat: Grounded Large Vision-Language Model for Remote Sensing.** 27831-27840
- Linus Scheibenreif, Michael Mommert, Damian Borth: **Parameter Efficient Self-Supervised Geospatial Domain Adaptation.** 27841-27851
- Boran Han, Shuai Zhang, Xingjian Shi, Markus Reichstein: **Bridging Remote Sensors with Multisensor Geospatial Foundation Models.** 27852-27862
- Lianggangxu Chen, Xuejiao Wang, Jiale Lu, Shaohui Lin, Changbo Wang, Gaoqi He: **CLIP-Driven Open-Vocabulary 3D Scene Graph Generation via Cross-Modality Contrastive Learning.** 27863-27873
- Romain Loiseau, Elliot Vincent, Mathieu Aubry, Loïc Landrieu: **Learnable Earth Parser: Discovering 3D Prototypes in Aerial Scans.** 27874-27884
- Xu Zheng, Pengyuan Zhou, Athanasios V. Vasilakos, Lin Wang: **Semantics, Distortion, and Style Matter: Towards Source-Free UDA for Panoramic Segmentation.** 27885-27895
- Guofeng Mei, Luigi Riz, Yiming Wang, Fabio Poiesi: **Geometrically-Driven Aggregation for Zero-Shot 3D Point Cloud Understanding.** 27896-27905
- Jiehong Lin, Lihua Liu, Dekun Lu, Kui Jia: **SAM-6D: Segment Anything Model Meets Zero-Shot 6D Object Pose Estimation.** 27906-27916
- Guangrui Li: **Construct to Associate: Cooperative Context Learning for Domain Adaptive Point Cloud Segmentation.** 27917-27926
- Yuqi Yang, Peng-Tao Jiang, Qibin Hou, Hao Zhang, Jinwei Chen, Bo Li: **Multi-Task Dense Prediction via Mixture of Low-Rank Experts.** 27927-27937
- Guan Wang, Zhimin Li, Qingchao Chen, Yang Liu: **OED: Towards One-stage End-to-End Dynamic Scene Graph Generation.** 27938-27947
- Xiangtai Li, Haobo Yuan, Wei Li, Henghui Ding, Size Wu, Wenwei Zhang, Yining Li, Kai Chen, Chen Change Loy: **OMG-Seg: Is One Model Good Enough for all Segmentation?** 27948-27959
- Hanrong Ye, Dan Xu: **DiffusionMTL: Learning Multi-Task Denoising Diffusion Model from Partially Annotated Data.** 27960-27969



- Guangzhi Wang, Yangyang Guo, Ziwei Xu, Mohan S. Kankanhalli: **Bilateral Adaptation for Human-Object Interaction Detection with Occlusion-Robustness.** 27970-27980
- Colton Stearns, Alex Fu, Jiateng Liu, Jeong Joon Park, Davis Rempe, Despoina Paschalidou, Leonidas J. Guibas: **CurveCloudNet: Processing Point Clouds with 1D Structure.** 27981-27991
- Jitesh Jain, Jianwei Yang, Humphrey Shi: **VCoder: Versatile Vision Encoders for Multimodal Large Language Models.** 27992-28002
- Guanqi Zhan, Chuanxia Zheng, Weidi Xie, Andrew Zisserman: **Amodal Ground Truth and Completion in the Wild.** 28003-28013
- Liyuan Zhu, Shengyu Huang, Konrad Schindler, Iro Armeni: **Living Scenes: Multi-object Relocalization and Reconstruction in Changing 3D Environments.** 28014-28024
- Zhuoxuan Peng, S.-H. Gary Chan: **Single Domain Generalization for Crowd Counting.** 28025-28034
- Jiaheng Liu, Jianhao Li, Kaisiyuan Wang, Hongcheng Guo, Jian Yang, Junran Peng, Ke Xu, Xianglong Liu, Jinyang Guo: **LTA-PCS: Learnable Task-Agnostic Point Cloud Sampling.** 28035-28045
- Xiaohong Zhang, Huisheng Ye, Jingwen Li, Qinyu Tang, Yuanqi Li, Yanwen Guo, Jie Guo: **Prompt3D: Random Prompt Assisted Weakly-Supervised 3D Object Detection.** 28046-28055
- Yu-Ju Tsai, Jin-Cheng Jhang, Jingjing Zheng, Wei Wang, Albert Y. C. Chen, Min Sun, Cheng-Hao Kuo, Ming-Hsuan Yang: **No More Ambiguity in 360° Room Layout via Bi-Layout Estimation.** 28056-28065
- Jinwon Ko, Dongkwon Jin, Chang-Su Kim: **Semantic Line Combination Detector.** 28066-28075
- Rongjie Li, Songyang Zhang, Dahua Lin, Kai Chen, Xuming He: **From Pixels to Graphs: Open-Vocabulary Scene Graph Generation with Vision-Language Models.** 28076-28086
- Yuan Dong, Chuan Fang, Liefeng Bo, Zilong Dong, Ping Tan: **PanoContext-Former: Panoramic Total Scene Understanding with a Transformer.** 28087-28097
- Gianluca Scarpellini, Stefano Fiorini, Francesco Giuliani, Pietro Morerio, Alessio Del Bue: **DiffAssemble: A Unified Graph-Diffusion Model for 2D and 3D Reassembly.** 28098-28108
- Yawen Lu, Dongfang Liu, Qifan Wang, Cheng Han, Yiming Cui, Zhiwen Cao, Xueling Zhang, Yingjie Victor Chen, Heng Fan : **ProMotion: Prototypes as Motion Learners.** 28109-28119
- Yichen Yao, Zimo Jiang, Yujing Sun, Zhencai Zhu, Xinge Zhu, Runnan Chen, Yuexin Ma: **HUNTER: Unsupervised Human-Centric 3D Detection via Transferring Knowledge from Synthetic Instances to Real Scenes.** 28120-28129
- Chuangchuang Tan, Huan Liu, Yao Zhao, Shikui Wei, Guanghua Gu, Ping Liu, Yunchao Wei: **Rethinking the Up-Sampling Operations in CNN-Based Generative Network for Generalizable Deepfake Detection.** 28130-28139
- Ayush Sarkar, Hanlin Mai, Amitabh Mahapatra, Svetlana Lazebnik, David A. Forsyth, Anand Bhattacharjee: **Shadows Don't Lie and Lines Can't Bend! Generative Models Don't know Projective Geometry...for Now.** 28140-28149
- Tianci Bi, Xiaoyi Zhang, Zhizheng Zhang, Wenxuan Xie, Cuiling Lan, Yan Lu, Nanning Zheng: **Text Grouping Adapter: Adapting Pre-Trained Text Detector for Layout Analysis.** 28150-28159
- Jongha Kim, Jihwan Park, Jinyoung Park, Jinyoung Kim, Sehyung Kim, Hyunwoo J. Kim: **Groupwise Query Specialization and Quality-Aware Multi-Assignment for Transformer-Based Visual Relationship Detection.** 28160-28169
- Ziqiang Zheng, Haixin Liang, Binh-Son Hua, Yue Him Wong, Put Ang, Apple Pui Yi Chui, Sai-Kit Yeung: **CoralSCOP: Segment any Coral Image on this Planet.** 28170-28180
- Huimin Huang, Yawen Huang, Lanfen Lin, Ruofeng Tong, Yen-Wei Chen, Hao Zheng, Yuexiang Li, Yefeng Zheng: **Going Beyond Multi-Task Dense Prediction with Synergy Embedding Models.** 28181-28190
- Zhuolong Li, Xingao Li, Changxing Ding, Xiangmin Xu: **Disentangled Pre-Training for Human-Object Interaction Detection.** 28191-28201
- Yuqian Yuan, Wentong Li, Jian Liu, Dongqi Tang, Xinjie Luo, Chi Qin, Lei Zhang, Jianke Zhu: **Osprey: Pixel Understanding with Visual Instruction Tuning.** 28202-28211
- Jinguo Luo, Weihong Ren, Weibo Jiang, Xi'ai Chen, Qiang Wang, Zhi Han, Honghai Liu: **Discovering Syntactic Interaction Clues for Human-Object Interaction Detection.** 28212-28222
- Simon Weber, Baris Zöngür, Nikita Araslanov, Daniel Cremers: **Flattening the Parent Bias: Hierarchical Semantic Segmentation in the Poincaré Ball.** 28223-28232
- Ce Zhang, Simon Stepputtis, Joseph Campbell, Katia P. Sycara, Yaqi Xie: **HiKER-SGG: Hierarchical Knowledge Enhanced Robust Scene Graph Generation.** 28233-28243
- Xin Kang, Lei Chu, Jiahao Li, Xuejin Chen, Yan Lu: **Hierarchical Intra-Modal Correlation Learning for Label-Free 3D Semantic Segmentation.** 28244-28253
- Zhihai Zhang, Jian Ding, Li Jiang, Dengxin Dai, Guisong Xia: **FreePoint: Unsupervised Point Cloud Instance Segmentation.** 28254-28263



- Weiming Zhang, Yexin Liu, Xu Zheng, Lin Wang:
GoodSAM: Bridging Domain and Capacity Gaps via Segment Anything Model for Distortion-Aware Panoramic Semantic Segmentation. 28264-28273
- Mi Yan, Jiazhao Zhang, Yan Zhu, He Wang:
MaskClustering: View Consensus Based Mask Graph Clustering for Open-Vocabulary 3D Instance Segmentation. 28274-28284
- Suraj Patni, Aradhya Agarwal, Chetan Arora:
ECoDepth: Effective Conditioning of Diffusion Models for Monocular Depth Estimation. 28285-28295
- Albert J. Zhai, Yuan Shen, Emily Y. Chen, Gloria X. Wang, Xinlei Wang, Sheng Wang, Kaiyu Guan, Shenlong Wang:
Physical Property Understanding from Language-Embedded Feature Fields. 28296-28305
- Kibum Kim, Kanghoon Yoon, Jaehyeong Jeon, Yeonjun In, Jinyoung Moon, Donghyun Kim, Chanyoung Park:
LLM4SGG: Large Language Models for Weakly Supervised Scene Graph Generation. 28306-28316
- Zeeshan Hayder, Xuming He:
DSGG: Dense Relation Transformer for an End-to-End Scene Graph Generation. 28317-28326
- Jianjun Xu, Yuxin Wang, Hongtao Xie, Yongdong Zhang:
OTE: Exploring Accurate Scene Text Recognition Using One Token. 28327-28336
- Jumin Lee, Sebin Lee, Changho Jo, Woobin Im, Juhyeong Seon, Sung-Eui Yoon:
SemCity: Semantic Scene Generation with Triplane Diffusion. 28337-28347
- Bowen Deng, Siyang Song, Andrew P. French, Denis Schluppeck, Michael P. Pound:
Advancing Saliency Ranking with Human Fixations: Dataset, Models and Benchmarks. 28348-28357
- Boqiang Zhang, Hongtao Xie, Zuan Gao, Yuxin Wang:
Choose What You Need: Disentangled Representation Learning for Scene Text Recognition, Removal and Editing. 28358-28368
- Jiankai Li, Yunhong Wang, Xiefan Guo, Ruijie Yang, Weixin Li:
Leveraging Predicate and Triplet Learning for Scene Graph Generation. 28369-28379
- Mingyue Guo, Li Yuan, Zhaoyi Yan, Binghui Chen, Yaowei Wang, Qixiang Ye:
Regressor-Segmenter Mutual Prompt Learning for Crowd Counting. 28380-28389
- Yuchen Zhou, Linkai Liu, Chao Gou:
Learning from Observer Gaze: Zero-Shot Attention Prediction Oriented by Human-Object Interaction Recognition. 28390-28400
- Yaxu Xie, Alain Pagani, Didier Stricker:
SG-PGM: Partial Graph Matching Network with Semantic Geometric Fusion for 3D Scene Graph Alignment and its Downstream Tasks. 28401-28411
- Xiangheng Shan, Dongyue Wu, Guilin Zhu, Yuanjie Shao, Nong Sang, Changxin Gao:
Open-Vocabulary Semantic Segmentation with Image Embedding Balancing. 28412-28421
- Aobo Li, Jinjian Wu, Yongxu Liu, Leida Li:
Bridging the Synthetic-to-Authentic Gap: Distortion-Guided Unsupervised Domain Adaptation for Blind Image Quality Assessment. 28422-28431
- Junhao Dong, Piotr Koniusz, Junxi Chen, Z. Jane Wang, Yew-Soon Ong:
Robust Distillation via Untargeted and Targeted Intermediate Adversarial Samples. 28432-28442
- Haitao Wen, Lili Pan, Yu Dai, Heqian Qiu, Lanxiao Wang, Qingbo Wu, Hongliang Li:
Class Incremental Learning with Multi-Teacher Distillation. 28443-28452
- Zhaoheng Zheng, Jingmin Wei, Xuefeng Hu, Haidong Zhu, Ram Nevatia:
Large Language Models are Good Prompt Learners for Low-Shot Image Classification. 28453-28462
- Zhanxin Gao, Jun Cen, Xiaobin Chang:
Consistent Prompting for Rehearsal-Free Continual Learning. 28463-28473
- Sicong Shen, Yang Zhou, Bingzheng Wei, Eric I-Chao Chang, Yan Xu:
Tuning Stable Rank Shrinkage: Aiming at the Overlooked Structural Risk in Fine-tuning. 28474-28484
- Guodong Ding, Hans Golong, Angela Yao:
Coherent Temporal Synthesis for Incremental Action Segmentation. 28485-28494
- Qiwei Li, Yuxin Peng, Jiahuan Zhou:
FCS: Feature Calibration and Separation for Non-Exemplar Class Incremental Learning. 28495-28504
- Shuai Shao, Yu Bai, Yan Wang, Baodi Liu, Yicong Zhou:
DELL: Direct-and-Inverse CLIP for Open-World Few-Shot Learning. 28505-28514
- Yu Mitsuzumi, Akisato Kimura, Hisashi Kashima:
Understanding and Improving Source-Free Domain Adaptation from a Theoretical Perspective. 28515-28524
- Dipam Goswami, Albin Soutif-Cormerais, Yuyang Liu, Sandesh Kamath, Bartłomiej Twardowski, Joost van de Weijer:
Resurrecting Old Classes with New Data for Exemplar-Free Continual Learning. 28525-28534
- Junhao Dong, Piotr Koniusz, Junxi Chen, Xiaohua Xie, Yew-Soon Ong:
Adversarially Robust Few-shot Learning via Parameter Co-distillation of Similarity and Class Concept Learners. 28535-28544



- Ba Hung Ngo, Nhat-Tuong Do-Tran, Tuan-Ngoc Nguyen, Hae-Gon Jeon, Tae Jong Choi: **Learning CNN on ViT: A Hybrid Model to Explicitly Class-Specific Boundaries for Domain Adaptation.** 28545-28554
- Haoyu He, Zizheng Pan, Jing Liu, Jianfei Cai, Bohan Zhuang: **Efficient Stitchable Task Adaptation.** 28555-28565
- Zhi Zhang, Qizhe Zhang, Zijun Gao, Renrui Zhang, Ekaterina Shutova, Shiji Zhou, Shanghang Zhang: **Gradient-based Parameter Selection for Efficient Fine-Tuning.** 28566-28577
- Xinyu Tian, Shu Zou, Zhao yuan Yang, Jing Zhang: **ArGue: Attribute-Guided Prompt Tuning for Vision-Language Models.** 28578-28587
- Hai Zhang, Junzhe Xu, Shanlin Jiang, Zhenan He: **Simple Semantic-Aided Few-Shot Learning.** 28588-28597
- Xi Wang, Xu Yang, Jie Yin, Kun Wei, Cheng Deng: **Long-Tail Class Incremental Learning via Independent SUb-Prototype Construction.** 28598-28607
- Guangxing Han, Ser-Nam Lim: **Few-Shot Object Detection with Foundation Models.** 28608-28618
- Zhixiang Wei, Lin Chen, Yi Jin, Xiaoxiao Ma, Tianle Liu, Pengyang Ling, Ben Wang, Huaian Chen, Jinjin Zheng: **Stronger, Fewer, & Superior: Harnessing Vision Foundation Models for Domain Generalized Semantic Segmentation.** 28619-28630
- Hongbo Zhao, Bolin Ni, Junsong Fan, Yuxi Wang, Yuntao Chen, Gaofeng Meng, Zhaoxiang Zhang: **Continual Forgetting for Pre-Trained Vision Models.** 28631-28642
- Taeckyung Lee, Sorn Chottananurak, Taesik Gong, Sung-Ju Lee: **AETTA: Label-Free Accuracy Estimation for Test-Time Adaptation.** 28643-28652
- Jiaming Liu, Ran Xu, Senqiao Yang, Renrui Zhang, Qizhe Zhang, Zehui Chen, Yandong Guo, Shanghang Zhang: **Continual-MAE: Adaptive Distribution Masked Autoencoders for Continual Test-Time Adaptation.** 28653-28663
- Zixuan Hu, Xiaotong Li, Shixiang Tang, Jun Liu, Yichun Hu, Ling-Yu Duan: **LEAD: Exploring Logit Space Evolution for Model Selection.** 28664-28673
- Minghao Fu, Ke Zhu: **Instance-based Max-margin for Practical Few-shot Recognition.** 28674-28683
- Yinong Oliver Wang, Younjoon Chung, Chen Henry Wu, Fernando De la Torre: **Domain Gap Embeddings for Generative Dataset Augmentation.** 28684-28694
- Yuncheng Guo, Xiaodong Gu: **JoAPR: Cleaning the Lens of Prompt Learning for Vision-Language Models.** 28695-28705
- Xusheng Cao, Haori Lu, Linlan Huang, Xialei Liu, Ming-Ming Cheng: **Generative Multi-modal Models are Good Class-Incremental Learners.** 28706-28717
- Yabin Zhang, Wenjie Zhu, Hui Tang, Zhiyuan Ma, Kaiyang Zhou, Lei Zhang: **Dual Memory Networks: A Versatile Adaptation Approach for Vision-Language Models.** 28718-28728
- Haiwen Diao, Bo Wan, Ying Zhang, Xu Jia, Huchuan Lu, Long Chen: **UniPT: Universal Parallel Tuning for Transfer Learning with Efficient Parameter and Memory.** 28729-28740
- Nan Pu, Wenjing Li, Xingyuan Ji, Yalan Qin, Nicu Sebe, Zhun Zhong: **Federated Generalized Category Discovery.** 28741-28750
- João Carreira, Michael King, Viorica Patraucean, Dilara Gokay, Catalin Ionescu, Yi Yang, Daniel Zoran, Joseph Heyward, Carl Doersch, Yusuf Aytar, Dima Damen, Andrew Zisserman: **Learning from One Continuous Video Stream.** 28751-28761
- Noor Ahmed, Anna Kukleva, Bernt Schiele: **OrCo: Towards Better Generalization via Orthogonality and Contrast for Few-Shot Class-Incremental Learning.** 28762-28771
- Junsu Kim, Hoseong Cho, Jihyeon Kim, Yihalem Yimolal Tiruneh, Seungryul Baek: **SDDGR: Stable Diffusion-Based Deep Generative Replay for Class Incremental Object Detection.** 28772-28781
- Yuzuru Nakamura, Yasunori Ishii, Takayoshi Yamashita: **Active Domain Adaptation with False Negative Prediction for Object Detection.** 28782-28792
- Niccolo Biondi, Federico Pernici, Simone Ricci, Alberto Del Bimbo: **Stationary Representations: Optimally Approximating Compatibility and Implications for Improved Model Replacements.** 28793-28804
- Ziming Hong, Li Shen, Tongliang Liu: **Your Transferability Barrier is Fragile: Free-Lunch for Transferring the Non-Transferable Learning.** 28805-28815
- Ségolène Martin, Yunshi Huang, Fereshteh Shakeri, Jean-Christophe Pesquet, Ismail Ben Ayed: **Transductive Zero-Shot and Few-Shot CLIP.** 28816-28826
- Rangel Daroya, Aaron Sun, Subhransu Maji: **Task2Box: Box Embeddings for Modeling Asymmetric Task Relationships.** 28827-28837
- Yajing Liu, Shijun Zhou, Xiyao Liu, Chunhui Hao, Baojie Fan, Jiandong Tian: **Unbiased Faster R-CNN for Single-source Domain Generalized Object Detection.** 28838-28847



last updated on 2024-12-11 20:32 CET by the dblp team



all metadata released as open data under CC0 1.0 license

See also: Terms of Use | Privacy Policy | Imprint

dblp was originally created in 1993 at:



since 2018, dblp has been operated and maintained by:



the dblp computer science bibliography is funded and supported by:

