Welcome to the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition in Seattle, Washington! CVPR is the premier and flagship annual meeting of IEEE/CVF and PAMI-TC, where researchers in our community present their latest advances in computer vision, pattern recognition, machine learning, robotics, and artificial intelligence, both in theory and practice. Our program includes invited keynote, oral and poster presentations, panels, tutorials, workshops, demos, exhibitions, and social events, all aimed at providing attendees with an exciting and enriching experience. CVPR 2024 is primarily an in-person conference, but for those who are unable to join us physically, we are pleased to offer a virtual component that will provide access to conference papers, posters, videos, and talks.

CVPR 2024 received 11,532 valid paper submissions, a 26% increase from CVPR 2023. The review process was managed by the 6 Program Co-Chairs, 24 Senior Area Chairs, and 477 Area Chairs. During the review phase, each paper received at least 3 reviews from a pool of 9,872 reviewers. As in prior years, after receiving these initial reviews, authors had the opportunity to submit a rebuttal to the reviews. The process concluded with discussion among reviewers and ACs, finalizing of reviews, and ACs working in triplets to make final accept/reject decisions for each paper. At the end of this process, 2,719 papers were accepted, for a 23.6% overall acceptance rate. In keeping with the CVPR tradition, the PCs did not pre-define any target acceptance rate or number of papers to be accepted; the resulting acceptance rate reflects the community consensus, and is consistent with past CVPRs.

All of the 2,719 accepted papers were invited to present posters at CVPR. In addition, 90 (3.3%) papers were selected to be presented as oral talks, based on nominations from Area Chairs, and 324 (11.9%) were selected by ACs to be "highlights" because of their high quality and potential impact. The highlights are flagged with a special annotation in the program. ACs nominated 24 papers to be best paper award candidates, from which a committee convened by the PCs selected the award winners. The award winners will be announced during the conference.

CVPR 2024 brings back the tradition of oral presentations in a three-track configuration. Nevertheless, we kept many of the innovations of CVPR 2023, including single-track panel discussions, "highlights" to indicate top-rated papers, the use of OpenReview for paper submission and management, and the role of Senior Area Chair to help oversee the review process.

We would like to thank everyone involved in making CVPR 2024 a success. This includes the organizing committee, the Senior Area Chairs, Area Chairs, and reviewers, the authors, the demo session participants, and the donors and exhibitors. David Forsyth’s service as Senior Advisor to the Program Chairs was incredibly helpful. We also thank Nicole Finn and her C to C Events team for organizing the conference logistics, Lee Campbell and the Event Hosts team for their work on the website and virtual platform, Mike Weil and Hall Erickson for handling sponsorships and the exhibition, and Luba Elliot as our inaugural AI Art Coordinator. Last but not least, we thank all of you for attending CVPR 2024 and making it one of the top venues for computer vision research in the world. We hope that you also have some time to explore Seattle during the conference.

Enjoy CVPR 2024. We look forward to meeting you in person!

Program Chairs
Ali Farhadi, (University of Washington)
David Crandall, (Indiana University)
Imari Sato, (National Institute of Informatics)
Jianxin Wu, (Nanjing University)
Robert Pless, (George Washington University)
Zeynep Akata, (University of Tübingen)

General Chairs
Octavia Camps, (Northeastern University)
Ramin Zabih, (Cornell University)
Rita Cucchiara, (University of Modena and Reggio Emilia)
Sudeep Sarkar, (University of South Florida)
Walter Scheirer, (University of Notre Dame)
CVPR 2024 ORGANIZING COMMITTEE

General Chairs
Octavia Camps (Northeastern University)
Ramin Zabih (Cornell)
Rita Cucchiara (University of Modena and Reggio Emilia)
Sudeep Sarkar (University of South Florida)
Walter Scheirer (University of Notre Dame)

Program Chairs
Ali Farhadi (University of Washington)
David Crandall (Indiana University)
Imari Sato (National Institute of Informatics)
Jianxin Wu (Nanjing University)
Robert Pless (George Washington University)
Zeynep Akata (University of Tübingen)

Senior Advisor to the Program Chairs
David Forsyth (University of Illinois at Urbana-Champaign)

Technical Chair
Yoshitomo Matsubara (Spiffy AI)

Tutorial Chairs
Katarina Doctor (U.S. Naval Research Lab)
Vitomir Struc (University of Ljubljana)

Workshop Chairs
Abhinav Shrivastava (University of Maryland)
Andrew Owens (University of Michigan)
Antitza Dantcheva (Inria)
Luisa Verdoliva (University Federico II of Naples)

DEI Chairs
Adriana Kovashka (Pitt)
CJ Taylor (Amazon AWS)
Michael King (Fiat)
Roni Sengupta (UNC)
Sara Beery (MIT)
Shuran Song (Columbia University)
Tamara L Berg

Conference Ombud
Angjoo Kanazawa (University of California Berkeley)
Derek Hoiem (University of Illinois at Urbana-Champaign)

Demonstration Chairs
Sathyanarayanan N. Aakur (Auburn University)
Shu Kong (University of Macau, Texas A&M University)

Senior PAMI-TC Ombud
David Forsyth (University of Illinois at Urbana-Champaign)
Linda Shapiro (UW Reality Lab University of Washington)

CVPR 2024 SENIOR AREA CHAIRS

Alex Schwing
Alexei A Efros
Aniruddha Kembhavi
Anthony Hoogs
Aude Oliva
Bernard Ghanem
Bharath Hariharan
Bohyung Han
Fernando De la Torre
Jingdong Wang
Jingyi Yu
Juan Carlos Niebles
Kyoung Mu Lee
Nathan Jacobs
Pascal Fua
Philipp Kraehenbuehl
Phillip Isola
Richard Souvenir
Roozbeh Mottaghi
Ruiyang Yang
Sing Bing Kang
Srinivasa Narasimhan
Tal Hassner
Yoichi Sato
We are grateful to all of the 9,872 reviewers who helped make CVPR 2024 possible. We are especially pleased to recognize the following Outstanding Reviewers, whose high-quality reviews (as judged by their Area Chairs) placed them among the top 2% of reviewers.

Aadarsh Sahoo
Aaron Walsman
Abhinav Shukla
Adnan Qayyum
Adrian Lopez
Adrian Penate-Sanchez
Ahmed Taha
Alessio Rompero
Alexander Hermans
Alexandros Haliassos
Alvaro Budria
Amine Bourki
Ana C Murillo
Anastasia Antsiferova
Andrea Conti
Andrew Brown
Anh-Quan Cao
Anil Usunuzbas
Anna Kukleva
Arjun Karpur
Armin Hadzic
Ashish Ramayee Asokan
Assia Benbhi
Ayush Jain
Ayush Saraf
Benjamin Busam
Bicheng Xu
Bingliang Jiao
Bhumiksha
Carlo Masone
Carlos Rodriguez-Pardo
Changqing Zhang
Chenglin Yang
Chengze Miaomiao Li
Cheston Tan
Chiao An Yang
Chi Xu
Chunghyun Park
Cristian Rodriguez-Opazo
Dahyun Kang
Daqi Liu
Wei Yang
Wei-Chih Hung
Weihong Deng
Weisong Shi
Wen Li
Wenwu Qian
Wenyi Ren
William Robson Schwartz
Xinlei Chen
Xiu-Shen Wei
Xueting Li
Xuming He
Xun Cao
Yale Song
Yan Huang
Yan Zhang
Yanchao Yang
Yang Song
YANG WANG
YANG YU
Yannis Kalantidis
Yaoao Liu
Yasuaki Makihara
Yebin Liu
Yedid Hoshen
Yen-Yu Lin
Yezhong Yang
Yu Wu
Yu-Hsuan Tsai
Yu-Ting Chen
Yubin Song
Yifei Huang
Yijun Li
Yin Li
Ying Fu
Ying Wu
Yinghuan Shi
Yingli Tian
Yinqiang Zheng
Yisen Wang
Yiyi Liao
Yonatan Bisk
Yong Jae Lee
Yosi Keller
Yu Cheng
Yu Kong
Yu Li
Yu Wu
Yu-Chuan Su
Yu-Xiong Wang
Yuanfeng Tian
Yuchao Dai
Yuki M Asano
Yulun Zhang
Yumin Suh
Yunchao Wei
Yung-Yu Chuang
Yunzhu Li
Yusuke Sugano
Yuyin Zhou
Zaid Harchaoui
Zhangyang Wang
Zhaowei Cai
Zhaoxiang Zhang
Zhe Gan
Zheng Zhang
Zhengming Ding
Zhengyi Li
Zhichong Wu
Zhizhong Han
Zhu Zhong
Ziwei Liu
Zongwei Zhou
Zsofi Kira
Zuxuan Wu
Zuzana Kukelova
Wednesday, June 19

**7:00 - 10:30** Oral Session 1A: Low-Level Vision
- Specularity Factorization for Low-Light Enhancement, Saurabh Saini, P. Narayanan
- FlowIE: Efficient Image Enhancement via Rectified Flow, Yixuan Zhu, Wenliang Zhao, Ao Li, Yansong Tang, Jie Zhou, Jiwen Lu
- Bilateral Event Mining and Complementary for Event Stream Super-Resolution, Zhilin Huang, Quanmin Liang, Yijie Yu, Chujun Qin, Xiawu Zheng, Kai Huang, Zikun Zhou, Wenming Yang
- FMA-Net: Flow-Guided Dynamic Filtering and Iterative Feature Refinement with Multi-Attention for Joint Video Super-Resolution and Deblurring, Geunhyuk Youk, Ji Young Oh, Munchurl Kim

**9:00 - 10:30** Oral Session 1B: Vision and Graphics
- GPLDD3: Latent Diffusion of 3D Shape Generative Models by Enforcing Geometric and Physical Priors, Yuanyong Dong, Qi Zuo, Xiaodong Gu, Weihao Yuan, Zhengyi Zhao, Zilong Dong, Liefeng Bo, Qixing Huang
- Retrieval-Augmented Layout Transformer for Content-Aware Layout Generation, Daichi Horita, Naoto Inoue, Kofaro Kikuchi, Kota Yamaguchi, Kiyooharu Aizawa
- Eclipse: Disambiguating Illumination and Materials using Unintended Shadows, Dor Verbin, Ben Middichten, Peter Hedman, Jonathan T. Barron, Todd Zickler, Pratul P. Srinivasan
- DiffusionLight: Light Probes for Free by Painting a Chrome Ball, Pakpon Phongthawee, Worameth Chinchuthakun, Nontapath Sinsunthitth, Varun Jampani, Amit Raj, Pramook Khungurn, Supasorn Suwajanakorn

**9:00 - 10:30** Oral Session 1C: Humans: Face, Body, Pose, Gesture, Movement
- MultiPly: Reconstruction of Multiple People from Monocular Video in the Wild, Zeren Jiang, Chen Guo, Manuel Kaufmann, Tianjian Jiang, Julien Valentin, Otmar Hilliges, Jie Song
- URHand: Universal Relightable Hands, 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 R. Roche, Nathan Mankovich, Gustau Camps-Valls, Tolga Birdal

**10:00 - 10:30** Poster Setup (Arch 4E)

**10:30 - 12:00** Poster Session 1 & Exhibit Hall (Arch 4A-E)
- Highlight paper (check it out)
- Award candidate paper (see award sessions)
Bidirectional Autoregressive Diffusion Model for Dance Generation, Canyu Zhang, Youbao Tang, Ning Zhang, Rui-E Sun, Mei Han, Jing Xiao, Song Wang

High-Quality Facial Geometry and Appearance Capture at Home, Yuxuan Han, Junfeng Lyu, Feng Xu

Multiple View Geometry Transformers for 3D Human Pose Estimation, Ziwei Liao, Jiandong Huang, Chunyu Wang, Han Hu, Steven L. Waslander

PACER+: On-Demand Pedestrian Animation Controller in Driving Scenarios, Jingbo Wang, Zhengyi Luo, Ye Yuan, Yuxuan Li, Bo Dai

I'M HOI: Inertia-aware Monocular Capture of 3D Human-Object Interactions, Chengfeng Zhao, Jueze Zhang, Jishen Du, Ziwei Shan, Junye Wang, Jingyi Yu, Jingya Wang, Lan Xu

HAVE-FUN: Human Avatar Reconstruction from Few-Shot Unconstrained Images, Xihe Yang, Xingyu Chen, Daisheng Gao, Shaohui Wang, Xiaoguang Han, Baoyuan Wang

Can Language Beat Numerical Regression? Language-Based Multimodal Trajectory Prediction, Inhan Bae, Junoh Lee, Hae-Gon Jeon

3D Human Pose Perception from Egocentric Stereo Videos, Hiroyasu Akada, Jang Wang, Vladislav Golyanik, Christian Theobalt

Egocentric Whole-Body Motion Capture with FisheyeVIT and Diffusion-Based Motion Refinement, Jian Wang, Zhe Cao, Diogo Luvizong, Lingjie Liu, Kripasindhu Sarkar, Danhang Tang, Thabo Beeler, Christian Theobalt

Human Gaussian Splitting: Real-time Rendering of Animatable Avatars, Arthur Moreau, Jiefei Song, Helisa Dhama, Richard Shaw, Yiren Zhou, Eduardo Pérez-Pellitero

OHTA: One-shot Hand Avatar via Data-driven Implicit Prior, Xiaozheng Zheng, Chao Wen, Zhuo Su, Zeran Xu, Zhaohui Li, Yang Zhao, Zhaohui Xue

HOIAnimator: Generating Text-prompt Human-object Animations using Novel Perceptive Diffusion Models, Wenfeng Song, Xinyu Zhang, Shuai Li, Yang Gao, Aimin Hao, Xia Hou, Chengzhizao Chen, Ning Li, Hong Qin

Arbitrary Motion Style Transfer with Multi-condition Motion Latent Diffusion Model, Wenfeng Song, Xingliang Jin, Shuai Li, Chengzhizao Chen, Aimin Hao, Xia Hou, Ning Li, Hong Qin

Single-View Scene Point Cloud Human Grass Generation, Yan-Kang Wang, Chengyi Xing, Yi-Lin Wei, Xiao-Ming Wu, Wei-Shi Zheng

Attention-Propagation Network for Egocentric Heatmap to 3D Pose Lifting, Taeho Kang, Youngki Lee


AnySkill: Learning Open-Vocabulary Physical Skill for Interactive Agents, Jiemin Ci, Tangyu Li, Nian Liu, Yaodong Yang, Yixin Zhu, Siyuan Huang

From a Bird's Eye View to See: Joint Camera and Subject Registration without the Camera Calibration, Zekun Qian, Ruize Han, Wei Feng, Song Wang

HMDC-Poser: On-Device real-time Human Motion Tracking from Scalable Sparse Observations, Peng Dai, Yang Zhang, Tao Liu, Zhen Fan, Tianyuan Du, Zhuo Su, Xiaozheng Zheng, Zeming Li

Monocular Identity-Conditioned Facial Reflectance Reconstruction, Xingyu Ren, Jiankang Deng, Yuhao Cheng, Jia Guo, Chao Ma, Yichao Yan, Wenhuan Zhu, Xiaokang Yang

GAvatar: Animatable 3D Gaussian Avatars with Implicit Mesh Learning, Ye Yuan, Xuelei Li, Yangyi Huang, Shalini De Mello, Koki Nagano, Jan Kautz, Umar Iqbal

Score-Guided Diffusion for 3D Human Recovery, Anastasios Stathopoulos, Ligang Han, Dimitris Metaxas

3D-Aware Face Editing using Warping-Guided Latent Direction Learning, Yuhao Cheng, Zuo Han, Xingyu Ren, Wenhuan Zhu, Zenghai Xu, Di Xu, Changpeng Yang, Yichao Yan

WANDR: Intention-guided Human Motion Generation, Markos Diomataris, Nikos Athenasiou, Omid Taheri, Xi Wang, Otmar Hilg説es, Michael J. Black
98
95
89
87
85
84
82
80
 Weinstein
79
78
77
76
weeney
68
65
62
58
55
52
49
46
43
40
37
34
31
28
25
22
19
16
13
10
7
4
1

Exploring Vision Transformers for 3D Human Motion-Language Models with Motion Patches, Qing Yu, Mikhij Tanaka, Kent Fujwara
85
NIFTY: Neural Object Interaction Fields for Guided Human Motion Synthesis, Nilesh Kulkarni, Davis Rempe, Kyle Genova, Abhijit Kundu, Justin Johnson, David Fouhey, Leonidas Guibas
86
DreamAvatar: Text-and-Shape Guided 3D Human Avatar Generation via Diffusion Models, Yukang Cao, Yan-Pei Cao, Kai Hai, Yang Shan, Kwan-Yee K. Wong
87
Person-in-WiFi 3D: End-to-End Multi-Person 3D Pose Estimation with Wi-Fi, Kangwei Yan, Fei Wang, Bo Qian, Han Ding, Jingsong Han, Xing Wei
88
ScoreHypo: Probabilistic Human Mesh Estimation with Hypothesis Scoring, Yuan Xu, Xiaoxuan Ma, Jiajun Su, Wentao Zhu, Yu Qiao, Yizhou Wang
89
Relightable and Animatable Neural Avatar from Sparse-View Video, Zhen Xu, Sida Peng, Chen Geng, Linzhan Mou, Zhihan Yan, Jiaming Sun, Huijun Bao, Xiaowei Zhou
90
Relightable Gaussian Codec Avatars, Shunsuke Saito, Gabriel Schwartz, Tomas Simon, Junxuan Li, Giljoo Nam
91
Audio to Photoreal Embodiment: Synthesizing Humans in Conversations, Evonne Ng, Javier Romero, Timur Bagautdinov, Shaojie Bai, Trevor Darrell, Angjoo Kanazawa, Alexander Richard
92
Closely Interactive Human Reconstruction with Proxemics and Physics-Guided Adaption, Buzhen Huang, Chen Li, Chongyang Xu, Liang Pan, Yangang Wang, Gim Hee Lee
93
Video-Based Human Pose Regression via Decoupled Space-Time Aggregation, Jihe Je, Wenwu Yang
94
Rethinking Generalizable Face Anti-spoofing via Hierarchical Prototype-guided Distribution Refinement in Hyperbolic Space, Chengyang Hu, Ke-Yue Zhang, Taiping Yao, Shouhong Ding, Lihuang Ma
95
MoML: Online Meta Adaptation for 3D Human Motion Prediction, Xiaoning Sun, Huaijiang Sun, Bin Li, Dong Wei, Wei Qin Li, Jianfeng Lu
96
KITRO: Refining Human Mesh by 2D Clues and Kinematic-tree Rotation, Fangyuan Yang, Kerui Gu, Angela Yao
97
Guess The Unseen: Dynamic 3D Scene Reconstruction from Partial 2D Glimpses, Inhee Lee, Byungjun Kim, Hanbyul Jou
98
PEGASUS: Personalized Generative 3D Avatars with Composable Attributes, Hyunsoo Cha, Byungjun Kim, Hanbyul Jou
99
Semantic Human Mesh Reconstruction with Textures, Xiaoyu Zhan, Jianxin Yang, Yuanqi Li, Jie Guo, Yanwen Guo, Wenping Wang
100
SDPose: Tokenized Pose Estimation via Circulation-Guide Self-Distillation, Sichen Chen, Yingyi Zhang, Siming Huang, Ran Yi, Ke Fan, Ruixin Zhang, Peixian Chen, Jun Wang, Shouhong Ding, Lihuang Ma
101
Mocap Everyone Everywhere: Lightweight Motion Capture With Smartwatches and a Head-Mounted Camera, Jiye Lee, Hanbyul Jou
102
DPMesh: Exploiting Diffusion Prior for occluded Human Mesh Recovery, Yixuan Zhu, Ao Li, Vansong Tang, Weiliang Zhao, Jie Zhou, Ji Zhou, Ji Zhou
103
DPHMs: Diffusion Parametric Head Models for Depth-based Tracking, Jiapeng Tang, Angela Dai, Yin Yu Nie, Lev Markhasin, Justus Thies, Matthias Nießner
104
KTPFormer: Kinematics and Trajectory Prior Knowledge-Enhanced Transformer for 3D Human Pose Estimation, Jiuhua Peng, Yangzong Zhong, P. Y. Mok
105
Exploiting Style Latent Flows for Generalizing Deepfake Video Detection, Jongwook Choi, Taehoon Kim, Yonghyun Jeong, Seungryu Baek, Jongwon Choi
106
EMAGE: Towards Unified Holistic Co-Speech Gesture Generation via Expressive Masked Audio Gesture Modeling, Haiyang Liu, Zhaoyu Zhu, Giorgio Becherini, Yichen Peng, Mingyang Su, You Zhou, Xuefei Zhe, Naoyu Iwamoto, Bo Zheng, Michael J. Black
107
A Unified Framework for Human-centric Point Cloud Video Understanding, Yiteng Xu, Kecheng Ye, Xiao Han, Yiming Ren, Xinge Zhu, Yuexin Ma
108
109
CLOAF: CoLisiOn-Aware Human Flow, Andrey Davydov, Martin Engelbrecht, Mathieu Salzmann, Pascal Fua
110
EventEccentric: 3D Human Motion Capture from Egocentric Event Streams, Christen MillerDurai, Hiroyasu Akada, Jian Wang, Diogo Luvizon, Christian Theobalt, Vladislav Golyanik
111
A Call to Reflect on Evaluation Practices for Age Estimation: Comparative Analysis of the State-of-the-Art and a Unified Benchmark, Jakub Paplíhám, Vojtěch Franc
112
Holopversions: Real-time Free-viewpoint Rendering of Humans from Sparse RGB Cameras, Ashwath Shetty, Marc Habermann, Guoxing Sun, Diogo Luvizon, Vladislav Golyanik, Christian Theobalt
113
Synergistic Global-space Camera and Human Reconstruction from Videos, Yizhou Zhao, Tuanteng Yang Wang, Bhiksha Raj, Min Xu, Jimei Yang, Chun-Hao Paul Huang
114
3D Face Tracking from 2D Video through Iterative Dense UV to Image Flow, Felix Taubner, Prashant Raina, Mathieu Tuli, Eu Wern Teh, Chul Lee, Jiming Xiang
115
UltrAvatar: A Realistic Animatable 3D Avatar Diffusion Model with Authenticity Guided Textures, Mingyuan Zhou, Rakib Hyder, Ziwei Xuan, Guojun Qi
116
OmniMotionGPT: Animal Motion Generation with Limited Data, Zhangshihao Yang, Mingyang Zhou, Mengyi Shan, Bingbing Wen, Ziwei Xuan, Mitch Hill, Junjie Bai, Guo-Jun Qi, Yalin Wang
117
Text-Guided 3D Face Synthesis - From Generation to Editing, Yunjie Wu, Yaping Meng, Zhipeng Hu, Linchong Li, Haoqian Wu, Kun Zhou, Weiwai Xu, Xin Yu
118
Multi-scale Dynamic and Hierarchical Relationship Modeling for Facial Action Units Recognition, Zihan Wang, Siyang Song, Cheng Luo, Songhe Deng, Weicheng Xie, Linlin Shen
119
LiveHPS: LiDAR-based Scene-level Human Pose and Shape Estimation in Free Environment, Yiming Ren, Xiao Han, Chengfeng Zhao, Jingya Wang, Lan Xu, Jingyi Yu, Yuxin Ma
120
FaceChain-Imagined: Freely Crafting High-Fidelity Diverse Talking Faces from Disentangled Audio, Chao Xu, Yang Liu, Jiazhen Xing, Weida Wang, Mingze Sun, Jun Dan, Tianxian Huang, Siyuang Li, Zhi-Qi Cheng, Ying Tai, Baoqiu Sun
121
OpticalDR: A Deep Optical Imaging Model for Privacy-Protective Depression Recognition, Yuchen Pan, Junjun Jiang, Kui Jiang, Zhihao Wu, Keyuan Yu, Xiangning Liu
122
SCMAE: Selective Correspondence Enhancement with Masked Autoencoder for Self-Supervised Landmark Estimation, Keja Yin, Varshantha Rao, Ruowei Jiang, Xudong Liu, Parham Aarabi, David B. Lindell
123
TokenHMR: Advancing Human Mesh Recovery with a Tokenized Pose Representation, Sai Kumar Dwivedi, Yu Sun, Priyanka Patel, Yao Feng, Michael J. Black
124
Optimizing Diffusion Noise Can Serve As Universal Motion Priors, Korrawe Karunatanakul, Konpat Preechakul, Emre Aksan, Thabo Beeler, Supasorn Suwajanakorn, Siyu Tang
125
M&M VTO: Multi-Garment Virtual Try-On and Editing, Luoyang Zhu, Yingwei Li, Nan Liu, Hao Peng, Dawei Yang, Ira Kemelmacher-Shlizerman
126
AvatarGPT: All-in-One Framework for Motion Understanding Planning Generation and Beyond, Zixiang Zhou, Yu Wan, Boayuan Wang
127
A Simple Baseline for Efficient Hand Mesh Reconstruction, Zhishan Zhou, Shihao Zhou, Zhi Lv, Minqiang Zou, Yao Tang, Jiajun Liang
128
VINECS: Video-based Neural Character Skinning, Zhouyicheng Liao, Vladislav Golyanik, Marc Habermann, Christian Theobalt
129
ConvoFusion: Multi-Modal Conversational Diffusion for Co-Speech Gesture Synthesis, Muhammad Hamza Mughal, Rishabh Dhabal, Ikhansul Habibie, Lucia Donatelli, Marc Habermann, Christian Theobalt
130
Programmable Motion Generation for Open-Set Motion Control Tasks, Hanchao Liu, Xiaohang Zhan, Shaoli Huang, Tai-Jiang Mu, Ying Shan
131
From Feature to Gaze: A Generalizable Replacement of Linear Layer for Gaze Estimation, Yiwei Bao, Feng Lu
172 Emotional Speech-driven 3D Body Animation via Disentangled Latent Diffusion, Kiran Chhatre, Radek Daněček, Nikos Athanasiou, Giorgio Becherini, Christopher Peters, Michael J. Black, Timo Bolkart

173 ProxyCap: Real-time Monocular Full-body Capture in World Space via Human-Centric Proxy-to-Motion Learning, Yuxiang Zhang, Hongwen Zhang, Liangxiao Hu, Jiajun Zhang, Hongwei Yi, Shengping Zhang, Yebin Liu

174 MAS: Multi-view Ancestral Sampling for 3D Motion Generation Using 2D Diffusion, Roy Kapon, Guy Tevet, Daniel Cohen-Or, Amit H. Bermano

175 Efficient 3D Implicit Head Avatar with Mesh-anchored Hash Table Blendingshapes, Qizian Bai, Feitong Tan, Sean Fanello, Rohit Pandey, Mingsong Dou, Shichen Liu, Ping Tan, Yinda Zhang

176 Neural Sign Actors: A Diffusion Model for 3D Sign Language Production from Text, Vasileios Baltatzis, Rolando Los Alexandros Potamias, Evangelos Ververas, Guoxiong Sun, Jiankang Deng, Stefanos Zafeiriou

177 RAM-Avatar: Real-time Photo-Realistic Avatar from Monocular Videos with Full-body Control, Xiang Deng, Zerong Zheng, Yuxiang Zhang, Jinning Sun, Chao Xu, Xiaodong Yang, Lizhen Wang, Yebin Liu

178 Sharingan: A Transformer Architecture for Multi-Person Gaze Following, Sameh Tafas, Anshul Gupta, Jean-Marc Odobez

179 Degrees of Freedom Matter: Inferring Dynamics from Point Trajectories, Yan Zhang, Sergey Prokudin, Marko Mihajlovic, Qianli Ma, Siyu Tang

180 Authentic Hand Avatar from a Phone Scan via Universal Hand Model, Gyeongsik Moon, Weiwen Xu, Rohan Joshi, Chenglei Wu, Takaki Shioratori

181 UniHuman: A Unified Model For Editing Human Images in the Wild, Nanfan Li, Qing Liu, Krishna Kumar Singh, Yilin Wang, Jianming Zhang, Bryan A. Plummer, Zhe Lin

182 BlockGCN: Redefine Topology Awareness for Skeleton-Based Action Recognition, Yunxuan Zhou, Xudong Yan, Zhi-Qi Cheng, Yan Yang, Qi Dai, Xian-Sheng Hua

183 GoMAvatar: Efficient Animatable Human Modeling from Monocular Video Using Gaussians-on-mesh, Jing Wen, Xiaofei Wu, Songcen Xu, Zhensong Zhang, Yue Wang, Rong Xiong

184 WHAM: Reconstructing World-grounded Humans with Accurate 3D Motion, Soyong Shin, Juyong Kim, Eni Halilaj, Michael J. Black

185 Self-Supervised Facial Representation Learning with Facial Region Awareness, Zheng Gao, Ioannis Patras

186 ChatPose: Chatting about 3D Human Pose, Yao Feng, Jing Lin, Sai Kumar Dwivedi, Yu Sun, Priyanka Patel, Michael J. Black

187 AUEditNet: Dual-Branch Facial Action Unit Intensity Manipulation with Implicit Disentanglement, Shiwei Jin, Zhen Wang, Lei Wang, Peng Liu, Ning Bi, Truong Nguyen

188 Towards a Simultaneous and Granular Identity-Expression Control in Personalized Face Generation, Renshuai Liu, Bowen Ma, Wei Zhang, Zhipeng Hu, Changjie Fan, Tingjie Liu, Yu Ding, Xuan Cheng

189 Pose2RM: Enhance 3D Human Pose Estimation on Unseen Camera Settings via Invariant Risk Minimization, Yanlu Cai, Weizhong Zhang, Yuan Wu, Cheng Jin

190 Rethinking Human Motion Prediction with Symplctic Integral, Haipeng Chen, Kedi Lyu, Zhengguang Liu, Yifang Yin, Xun Yang, Yingda Lyu

191 Multimodal Sense-Informed Forecasting of 3D Human Motions, Zhiyong Wang, Qiucheng Li, Hengfang Wang, Xu Tang, Hong Zhou

192 Semantics-aware Motion Retargeting with Vision-Language Models, Haozong Zhang, Zhike Chen, Haocheng Xu, Lei Hao, Xiaofei Wu, Songcen Xu, Zhenzong Zhang, Yue Wang, Rong Xiong

193 Makeup Prior Models for 3D Facial Makeup Estimation and Applications, Xizhao Yang, Takafumi Taketomi, Yuki Endo, Yoshihiro Kanamori

194 FaceCom: Towards High-fidelity 3D Facial Shape Completion via Optimization and Inpainting Guidance, Yinglong Li, Hongyu Wu, Xiaogang Wang, Qingzhao Qin, Yijiao Zhao, Yong Wang, Aimin Hao

195 When StyleGAN Meets Stable Diffusion: a W+ Adapter for Personalized Image Generation, Xiaoming Li, Xinyu Hou, Chen Change Loy

196 MANUS: Markless Grasp Capture using Articulated 3D Gaussians, Chandradeep Pokhriya, Isaha Nikhil Shah, Angela Xing, Zejun Li, Kefan Chen, Avinash Sharma, Srinath Sridhar

197 Loose Inertial Poser: Motion Capture with IMU-attached Loose-Wear Jacket, Chengxu Zuo, Yiming Wang, Lishuang Zhan, Shihui Guo, Xinyu Yi, Feng Xu, Yipeng Qin

198 Anatomically Constrained Implicit Face Models, Prashanth Chandran, Gaspard Zoss

199 DiffusionRegPose: Enhancing Multi-Person Pose Estimation using a Diffusion-Based End-to-End Regression Approach, Dayi Tan, Hansheng Chen, Wei Tian, Lu Xiong

200 A Dual-Augmentor Framework for Domain Generalization in 3D Human Pose Estimation, Quucheng Peng, Ce Zheng, Chen Chen

201 RELIHD: A Comprehensive Multimodal Human Motion Dataset and Method, Ming Yan, Yan Zhang, Shuchang Cai, Shuqi Fan, Xincheng Lin, Yudi Dai, Siqi Shen, Chenglu Wen, Lan Xu, Yuexin Ma, Cheng Wang

202 Co-Speech Gesture Video Generation via Motion-Decoupled Diffusion Model, Xu He, Qiaoqiu Huang, Zhensong Zhang, Zhiwei Lin, Zhiyong Wu, Sicheng Yang, Mingli Li, Zhiyi Chen, Songcen Xu, Xiaofei Wu

203 HandDiff: 3D Hand Pose Estimation with Diffusion on Image-Point Cloud, Wencan Cheng, Hao Tang, Luc Van Gool, Jiong Hwan Ko

204 Normalizing Flows on the Product Space of SO(3) Manifolds for Probabilistic Human Pose Modeling, Olaf Dünkel, Tim Salzmann, Florian Pfaff

205 Towards Robust 3D Pose Transfer with Adversarial Learning, Haoyu Chen, Hao Tang, Ehsan Adeli, Guoying Zhao

206 PhysPT: Physics-aware Pretrained Transformer for Estimating Human Dynamics from Monocular Videos, Yufei Zhang, Jeffrey O. Kephart, Zijun Cui, Qiang Ji

207 HumMUS: Human Motion Understanding using State Space Models, Arnab Mondal, Stefano Alletto, Denis Tome

208 MultiPhys: Multi-Person Physics-aware 3D Motion Estimation, Nicolas Ugrinovic, Boxian Pan, Georgios Pavlakos, Despoina Paschalidou, Bokui Shen, Jordi Sanchez-Riera, Francesc Moreno-Noguer, Leonidas Guibas

209 Physics-Aware Hand-Object Interaction Denoising, Haowen Luo, Yunze Liu, Li Yi

210 HOI-FORMER: Hand-held Objects Identification Segmentation and Tracking in the Wild, Supreeth Narasimhaswamy, Huy Anh Nguyen, Lihan Huang, Minh Hoai

211 SCULPT: Shape-Conditioned Unpaired Learning of Pose-dependent Clothed and Textured Human Meshes, Soubhik Sanyal, Partha Ghosh, Jinyong Yang, Michael J. Black, Justus Thies, Timo Bolkart

212 PFStorer: Personalized Face Restoration and Super-Resolution, Tuomas Varanka, Tapani Toivonen, Soumin Tripathy, Guoying Zhao, Erman Acar

213 MS-MANO: Enabling Hand Pose Tracking with Biomechanical Constraints, Pengfei Xie, Wenqiang Xu, Tutian Tang, Zhenjun Yu, Cewu Lu

214 BOTH2Hands: Inferring 3D Hands from Both Text Prompts and Body Dynamics, Wenzian Zhang, Molin Huang, Xuyuan Zhou, Juzhe Zhang, Jingyi Yu, Jingya Wang, Lan Xu

215 MeshPose: Unifying DensePose and 3D Body Mesh Reconstruction, Eric-Tuan Le, Antonis Kakolyris, Petros Koutras, Himmy Tam, Efstratios Skordos, George Papandreou, Riza Alp Güler, Iasonas Kokkinos

216 CustomListener: Text-guided Responsive Interaction for User-friendly Listening Head Generation, Recent Listening Head Generation, Xi Liu, Ying Guo, Cheng Zhen, Tong Li, Yingqian Ao, Pengfei Yan

217 Generalizable Face Landmarking Guided by Conditional Face Warping, Jiayi Liang, Haotian Liu, Hengtong Xu, Dixin Luo

218 Skeleton-in-context: Unified Skeleton Sequence Modeling with In-context Learning, Xinshun Wang, Zhongbin Fang, Xia Li, Xiangta Li, Chen Chen, Mengyuan Liu
270 DSL-FIQA: Assessing Facial Image Quality via Dual-Set Degradation Learning and Landmark-Guided Transformer, Wei-Ting Chen, Gurunandan Krishnan, Qiang Gao, Sy-Yen Kuo, Szihou Ma, Jian Wang

271 CLIPtone: Unsupervised Learning for Text-based Image Tone Adjustment, Hyeongmin Lee, Kyoungook Kang, Jungseul Ok, Sunghyun Cho

272 Adapt or Perish: Adaptive Sparse Transformer with Attentive Feature Refinement for Image Restoration, Shihao Zhou, Duosheng Chen, Jinshan Pan, Jinglei Shi, Jufeng Yang

273 CPQA: Coding Priors-Guided Aggregation Network for Compressed Video Quality Enhancement, Qiang Zhu, Jinhua Hao, Yukang Ding, Yu Liu, Qiao Mo, Ming Sun, Chao Zhou, Shuyuan Zhu

274 Learning to Control Camera Exposure via Reinforcement Learning, Kyunghyun Lee, Ukcheol Shin, Byeong-Ok Lee

275 Real-Time Exposure Correction via Collaborative Transformations and Adaptive Sampling, Zhiwen Li, Feng Zhang, Meng Cao, Jipu Zhang, Yuanjie Shao, Yuehua Wang, Nong Sang

276 Towards Progressive Multi-Frequency Representation for Image Warping, Jun Xiao, Zilang Lyu, Cong Zhang, Yakun Ju, Changjian Shui, Xin-Man Lam

277 HIR-Diff: Unsupervised Hyperspectral Image Restoration via Improved Diffusion Models, Li Pang, Xiangyu Rui, Long Cui, Hongzhong Wang, Deyu Meng, Xiangyong Cao

278 ZERO-IG: Zero-Shot Illumination-Guided Joint Denoising and Adaptive Enhancement for Low-Light Images, Yiqi Shi, Duo Liu, Liguang Ye Tian, Xuezi Xia, Xiaojing Fu

279 Masked and Shuffled Blind Spot Denoising for Real-World Images, Hamadi Chaiaoui, Paolo Favaro

280 Continuous Optical Zooming: A Benchmark for Arbitrary-Scale Image Super-Resolution in Real World, Huiyuan Fu, Fei Peng, Xianwei Li, Yequn Li, Xin Wang, Huadong Ma


282 SD2Event: Self-supervised Learning of Dynamic Detectors and Contextual Descriptors for Event Cameras, Yuan Gao, Yuqing Zhu, Xinzun Li, Yimin Du, Tianzhu Zhang

283 LiFS: When Large Language Models Meet Few-Shot Segmentation, Lanyn Zhu, Tianrun Chen, Deji Y, Jieping Ye, Jun Liu

284 Telling Left from Right: Identifying Geometry-Aware Semantic Correspondence, Junyi Zhang, Charles Herrmann, Junhua Wu, Eric Chen, Varun Jampani, Deqing Sun, Ming-hsuan Yang

285 One-Shot Open Affordance Learning with Foundation Models, Gen Li, Deqing Sun, Laura Sevilla-Lara, Varun Jampani

286 CorrMatch: Label Propagation via Correlation Matching for Semi-Supervised Semantic Segmentation, Boyuan Li, Yuqi Yang, Le Zhang, Ming-Ming Cheng, Qibin Hou

287 Collaborating Foundation Models for Domain Generalized Semantic Segmentation, Yasser Benigmim, Subhankar Roy, Slim Essid, Vicky Kalogeiton, Stéphane Lathuilière

288 FocSAM: Delving Deeply into Focused Objects in Segmenting Anything, You Huang, Zongyu Lan, Lijuan Cao, Xianming Lin, Yichen Li, Ziqian Yang, Tianhong Dai, Bingfeng Zhang, Jimin Xiao

289 Finlsler-Laplace-Beltrami Operators with Application to Shape Analysis, Simon Weber, Thomas Dagès, Malin Gao, Daniel Cremers

290 Neural Implicit Representation for Building Digital Twins of Unknown Articulated Objects, Yijia Weng, Bowen Wen, Jonathan Tremblay, Vahts Blusik, Dieter Foer, Leonidas Guibas, Stan Birchfield

291 Putting the Object Back into Video Object Segmentation, • Ho Kei Cheng, Seoung Wu Oh, Brian Price, Joon-Young Lee, Alexander Schwing

292 BA-SAM: Scalable Bias-Mode Attention Mask for Segmentation Anything Model, Yiran Song, Qianyu Zhou, Xiangta Li, Deng-Ping Fan, Xuequn Lu, Lizhuang Huang

293 Task-aligned Part-aware Panoptic Segmentation through Joint Object-Part Representations, Daan de Geus, Gis Dubbleman

294 Open-World Semantic Segmentation Including Class Similarity, Matteo Sodano, Federico Magistri, Lucas Nunes, Jens Behley, Cyrill Stachniss

295 Hierarchical Histogram Threshold Segmentation – Auto-terminating High-detail Oversegmentation, Thomas V. Chang, Simon Seibt, Bartosz von Rymon Lipinski

296 AlignSAM: Aligning Segment Anything Model to Open Context via Reinforcement Learning, Duojun Huang, Xinyu Xiong, Jie Ma, Jichang Li, Zequn Jie, Lin Ma, Guanbin Li

297 SANerf-HQ: Segment Anything for NeRF in High Quality, Yichen Liu, Benran Hu, Chi-Keung Tang, Yu-Wing Tai

298 UniVS: Unified and Universal Video Segmentation with Prompts as Queries, Minghan Li, Shuai Li, Xindong Zhang, Lei Zhang

299 RankED: Addressing Imbalance and Uncertainty in Edge Detection Using Ranking-based Losses, Bedrettin Cetinkaya, Sinan Kalkan, Emre Akbas

300 Event-assisted Low-Light Video Object Segmentation, Hebei Li, Jin Wang, Jiahui Yuan, Yue Li, Wenming Weng, Yansong Peng, Yueyi Zhang, Zhiwei Xiong, Xiaoyan Sun

301 Density-Guided Semi-Supervised 3D Semantic Segmentation with Dual-Space Hardness Sampling, Jianan Li, Qiuei Dong

302 Exploring Regional Clues in CLIP for Zero-Shot Semantic Segmentation, Yi Zhang, Meng-Hao Guo, Miaow Wang, Shi-Min Lin

303 Category-Level Multi-Part Multi-Joint 3D Shape Assembly, Yichen Li, Kaichun Mo, Yueqi Duan, He Wang, Jieqian Zhang, Lin Shao

304 SA13D: Segment Any Instance in 3D Scenes, Yingda Yin, Yuzheng Liu, Yang Xiao, Daniel Cohen-Or, Jingwei Huang, Baoquan Chen

305 Towards the Uncharted: Density-Descending Feature Perturbation for Semi-supervised Semantic Segmentation, Xiaoyang Wang, Huihui Dai, Limin Yu, Yao Zhao, Jinmin Xiao

306 Hybrid Functional Maps for Create-Aware Non-Isometric Shape Matching, Lennart Bastian, Yizheng Xie, Nassir Navab, Zorah Lähner

307 Hunting Attributes: Context Prototype-Aware Learning for Weakly Supervised Semantic Segmentation, Feilong Tang, Zhongxing Xu, Zhaojun Qu, Wei Feng, Xinjiang Jiang, Zongyuan Ge

308 Self-Calibrating Vicinal Patch Minimisation for Model Calibration, Jianwei Liu, Changkun Ye, Ruikai Cui, Nick Barnes

309 ECLIPSE: Efficient Continual Learning in Panoptic Segmentation with Visual Prompt Tuning, Beomyoung Kim, Joosang Yu, Sung Ju Hwang

310 Clustering Propagation for Universal Medical Image Segmentation, Yuhang Ding, Lielie Li, Wenguang Wang, Yi Yang

311 Addressing Background Context Bias in Few-Shot Segmentation through Iterative Modulation, Lanyn Zhu, Tianrun Chen, Jianxiong Yin, Simon See, Jun Liu

312 Cross-Domain Few-Shot Segmentation via Iterative Support-Query Correspondence Mining, Jiahao Nie, Yun Xing, Gongjie Zhang, Pei Yan, Aoran Xiao, Yap-Peng Tan, Alex C. Kot, Shijian Lu

313 RankMatch: Exploring the Better Consistency Regularization for Semi-supervised Semantic Segmentation, Huayu Mai, Rui Sun, Tianzhu Zhang, Feng Wu

314 QDFormer: Towards Robust Audiovisual Segmentation in Complex Environments with Quantization-based Semantic Decomposition, Xiang Li, Jinguo Wang, Xiaohao Xu, Xiliu Peng, Rita Singh, Yan Lu, Biksha Raj

315 Frequency-Adaptive Dilated Convolution for Semantic • Segmentation, Linwei Chen, Lin Gu, Dezhui Zheng, Ying Fu

316 SED: A Simple Encoder-Decoder for Open-Vocabulary Semantic Segmentation, Bin Xie, Jiale Cao, Jin Xie, Fahad Shahbaz Khan, Yanwei Pang

317 PSDPM: Prototype-based Secondary Discriminative Pixels Mining for Weakly Supervised Semantic Segmentation, Xinqiao Zhao, Ziqian Yang, Tianhong Dai, Bingfeng Zhang, Jinmin Xiao

318 Coupled Laplacian Eigenmaps for Locally-Aware 3D Rigid Point Cloud Matching, Matteo Basico, Etienne Decencière, Laurent Corté, Yannick Tillier, David Ryckelynck

319 Universal Segmentation at Arbitrary Granularity with Language Instruction, Yong Li, Caolong Zhang, Yifeng Wang, Jiahao Wang, Yujiu Yang, Yansong Tang

320 PartDistill: 3D Shape Part Segmentation by Vision-Language Model Distillation, Ardian Umam, Cheng-Kun Yang, Min-Hung Chen, Jen-Hui Chuang, Yen-Yu Lin
He Guo, Zixuan Ye, Zhiguo Cao, Hao Lu
Unsupervised Semantic Segmentation, Yian Zhao, Kehan Li, Zesen Cheng, Pengchong Qiao, Xiawu Zheng, Lingxi Xi, Qi Tian, Wei Shen

CADTalk: An Algorithm and Benchmark for Semantic

Commenting of CAD Programs, Haocheng Yuan, Jing Xu, Hao Pan, Adrien Bousseau, Niloy J. Mitra, Changjian Li

Point2CAD: Reverse Engineering CAD Models from 3D Point

Clouds, Yujia Liu, Anton Obukhov, Jan Dirk Wegner, Konrad Schindler

ReThinking Interactive Image Segmentation with Low Latency

High Quality and Diverse Prompts, Qin Liu, Jaemin Cho, Mohit Bansal, Marc Niethammer

General Object Foundation Model for Images and Videos at

Scale, Junfeng Wu, Yi Jiang, Qihao Liu, Zehuan Yuan, Xiang Bai, Song Bai

Frozen CLIP: A Strong Backbone for Weakly Supervised Semantic

Segmentation, Bingfeng Zhang, Siyue Yu, Yunchao Wei, Yao Zhao, Jimin Xiao

Guided Slot Attention for Unsupervised Video Object

Segmentation, Minhyeok Lee, Sukwon Cho, Dogyoon Lee, Chaewon Park, Jungho Lee, Sangyoun Lee

Unlocking the Potential of Pre-trained Vision Transformers for Few-Shot Semantic Segmentation through Relationship Descriptors, Qizin Zhou, Hai-Ming Xu, Yangyang Shu, Lingqiao Liu

Grounding Everything: Emerging Localization Properties in Vision-Language Transformers, Wald Bouselham, Felix Petersen, Vittorio Ferrari, Hilde Kuehne

No Time to Train: Empowering Non-Parametric Networks for

Few-shot 3D Scene Segmentation, Xiangyang Zhu, Renrui Zhang, Bowei He, Ziyou Guo, Jiaming Liu, Han Xiao, Chaoyou Fu, Hao Dong, Peng Gao

Continual Segmentation with Disentangled Objectness Learning and Class Recognition, Yizheng Gong, Siyue Yu, Xiaoyang Wang, Jimin Xiao

GSVA: Generalized Segmentation via Multimodal Large

Language Models, Zhuofan Xia, Dongchen Han, Yizeng Han, Xuran Pan, Shijii Song, Gao Huang

MaGGlle: Masked Guided General Human Instance

Matting, Chuong Huynh, Seoung Wug Oh, Abhinav Shivastava, Jooyoung Lee

Enhanced Transformer towards Semantic-Contour

Features of Foreground for Portraits Matting, Zitao Wang, Qiguang Miao, Yue Xi, Peipei Zhao

Segment Any Event Streams via Weighted Adaptation of Pivotal Tokens, Zhiwen Chen, Zhiyu Zhu, Yifan Zhang, Junhui Hou, Guangming Shu, Jinjian Wu

PolarMatte: Fully Computational Ground-Truth-Quality Alpha

Matte Extraction for Images and Video using Polarized Screen Matting, Kenji Enomoto, TJ Rhodes, Brian Price, Gavin Miller

Segment Every Out-of-Distribution Object, Wenjie Zhao, Jia Li, Xin Dong, Yu Xiang, Yunhui Guo

Multi-view Aggregation Network for Dichotomous Image

Segmentation, Qian Yu, Xiaojia Zhao, Yuweei Pang, Lihe Zhang, Zhiwu Liu

pix2gastalt: Amodal Segmentation by Synthesizing Wholes,

Ege Ozguroglu, Ruoshi Liu, Didac Suris, Dian Chen, Achal Dave, Pavel Tokmakov, Carl Vondrick

Rethinking Prior Information Generation with CLIP for Few-Shot

Segmentation, Jin Wang, Bingfeng Zhang, Jian Pang, Honglong Chen, Wei Feng Liu

Image-to-Image Matching via Foundation Models: A New Perspective for Open-Vocabulary Semantic Segmentation, Yuan Wang, Rui Sun, Naosong Luo, Yuwen Pan, Tianzhu Zhang

Domain Separation Graph Neural Networks for Salienty Object

Ranking, Zijian Wu, Jun Lu, Jing Han, Lianfa Bai, Yi Zhang, Zhuang Zhao, Siyang Song

DIOD: Self-Distillation Meets Object Discovery, Sandra Kara, Hejer Ammar, Julien Denize, Florian Chabot, Quoc-Cuong Pham

Rethinking Few-shot 3D Point Cloud Semantic Segmentation, Zhaohong An, Guolei Sun, Yun Liu, Fayao Liu, Zongwei Wu, Dan Wang, Lu Van Gool, Sergio Belongie

Training Vision Transformers for Semi-Supervised Semantic Segmentation, Xinting Li, Ji Liang, Bernt Schiele

Open3DLS: Open-Vocabulary 3D Instance Segmentation with 2D Mask Guidance, Phuc Nguyen, Tuan Duc Ngo, Evangelos Kalogerakis, Chuang Gan, Anh Tran, Cuong Pham, Khi Nguyen

Emergent Open-Vocabulary Semantic Segmentation from Off-the-shelf Vision-Language Models, Jiayun Luo, Siddhesh Khandelwal, Leonid Sigal, Boyang Li

Memory-Scalable and Simplified Functional Map Learning, Robin Magnett, Marks Ovsjanikov

MFP: Making Full Use of Probability Maps for Interactive Image Segmentation, Chaewon Lee, Seon-Ho Lee, Chang-Su Kim

Spherical Mask: Coarse-to-Fine 3D Point Cloud Instance Segmentation with Spherical Representation, Sangyun Shin, Kaichen Zhou, Madhu Vankadari, Andrew Markham, Niki Trigoni

Adaptive Bidirectional Displacement for Semi-Supervised Medical Image Segmentation, Hanyang Chi, Jian Pang, Bingfeng Zhang, Weifeng Liu

RobustSAM: Segment Anything Robustly on Degraded Images, Wei-Ting Chen, Yu-Jiet Vong, Sy-Yen Kuo, Sizhou Ma, Jian Wang

LAKE-RED: Camouflaged Images Generation by Latent Background Knowledge Retrieval-Augmented Diffusion, Pancheng Zhao, Peng Xu, Penda Qin, Deng-Ping Fan, Zhicheng Zhang, Guoli Jia, Bowen Zhou, Jufeng Yang

Learn to Rectify the Bias of CLIP for Unsupervised Semantic Segmentation, Jingyun Wang, Guoliang Kang

CAT-SEg: Cost Aggregation for Open-Vocabulary Semantic Segmentation, Seokju Cho, Heeseong Shin, Sungwhan Hong, Anurag Arnab, Paul Hongguck Seo, Seungryong Kim

Prompt-Driven Referring Image Segmentation with Instance Contrasting, Chaohong Shang, Zichen Song, Heqian Qi, Lanxiao Wang, Fanman Meng, Hongliang Li

Kandinsky Conformal Prediction: Efficient Calibration of Image Segmentation Algorithms, Joren Brunkeef, Eric Marcus, Ray Shoembarsong, Jan-Jakob Sonke, Jonas Teuwen

OVFoodSeg: Elevating Open-Vocabulary Food Image Segmentation via Image-Informed Textual Representation, Xiongwei Wu, Sicheng Yu, Ee-Peng Lim, Chong-Wah Ngo

Back to 3D: Few-Shot 3D Keypoint Detection with Back-Projected 2D Features, Thomas Wimmer, Peter Wonka, Maks Ovsjanikov

Deciphering ‘What’ and ‘Where’ Visual Pathways from Spectral Clustering of Layer-Distributed Neural Representations, Xiao Zhang, David Unis, Michael Mair

Open Vocabulary Semantic Scene Sketch Understanding, Ahmed Bourouis, Judith E. Fan, Yulia Gryaditskaya

USE: Universal Segment Embeddings for Open-Vocabulary Image Segmentation, Xiaoxi Wang, Wenbin He, Xiwei Xuan, Clint Sebastian, Jorge Piazzentino Ono, Xin Li, Sima Behpour, Thang Doan, Liang Gou, Han-Wei Shen, Liu Ren

Diff-Plugin: Revitalizing Details for Diffusion-based Low-level Tasks, Yuhao Liu, Zhanhao Ke, Fang Liu, Nanxuan Zhao, Rynson W.H. Lau

XCube: Large-Scale 3D Generative Modeling using Sparse Voxel Hierarchies, Xuanchi Ren, Jiahui Huang, Xiaohui Zeng, Ken Museth, Sanja Fidler, Francis Williams

SC-GS: Sparse-Controlled Gaussian Splitting for Editable Dynamic Scenes, Yi-Hua Huang, Yang-Tian Sun, Ziyi Yang, Xiaoyang Lyu, Yan-Pei Cao, Xiaojuan Qi

StyLitGAN: Image-Based Relighting via Latent Control, Anand Bhattad, James Soole, D.A. Forsyth

GPLD3D: Latent Diffusion of 3D Shape Generative Models by Enforcing Geometric and Physical Prioris, Yuan Dong, Qi Zuo, Xiaodong Gu, Weihao Yuan, Zhengyi Zhao, Zilong Dong, Liefeng Bo, Qixing Huang

Image Sculpting: Precise Object Editing with 3D Geometry Control, Jiraphon Yenphraphat, Xichen Fan, Sainan Liu, Daniele Panazzo, Saining Xie

Paint3D: Paint Anything 3D with Lighting-Less Texture Diffusion Models, Xianfang Zeng, Xin Chen, Zhongqi Qi, Wen Liu, Zibo Zhao, Zhibin Wang, Bin Fu, Yong Liu, Gang Yu

Retrieval-Augmented Layout Transformer for Content-Aware Layout Generation, Daichi Horita, Naoto Inoue, Kotaro Kikuchi, Kota Yamaguchi, Kiyoharu Aizawa

Holo-Relighting: Controllable Volumetric Portrait Relighting from a Single Image, Yiqin Mei, Yu Zeng, He Zhang, Zhixin Shu, Xuaner Zhang, Sai Bi, Jianming Zhang, HyunJoon Jung, Vishal M. Patel

Neural Fields as Distributions: Signal Processing Beyond Euclidean Space, Daniel Reban, Soroosh Yazdani, Kwang Moo Yi, Andrea Tagliasacchi

Eclipse: Disambiguating Illumination and Materials using Unintended Shadows, Dor Verbin, Ben Mildenhall, Peter Hedman, Jonathan T. Barron, Todd Zickler, Pratul P. Srinivasan

TexOct: Generating Textures of 3D Models with Octree-based Diffusion, Jialun Liu, Chenming Wu, Xiqin Liu, Xing Liu, Jinbo Wu, Haotian Peng, Chen Zhao, Haocheng Feng, Jingtuo Liu, Errui Ding

Differentiable Micro-Mesh Construction, Yishou Dou, Zhong Zheng, Qiaoqiao Jin, Rui Shi, Yuhan Li, Bingbing Ni

TextureDreamer: Image-Guided Texture Synthesis Through Geometry-Aware Diffusion, Yu-Ying Yeh, Jia-Bin Huang, Changli Kim, Lei Xiao, Thu Nguyen-Phuc, Numair Khan, Chen Zhang, Mannoman Chandraker, Carl S Marshall, Zhao Dong, Zhengxin Li

As-Plausible-As-Possible: Plausibility-Aware Mesh Deformation Using 2D Diffusion Priors, Seungwook Yoo, Kunho Kim, Vladimir G. Kim, Mihnyuk Sung

Breathing Life Into Sketches Using Text-to-Video Priors, Rinon Gal, Yael Vinker, Yuval Alaluf, Amit Bermano, Daniel Cohen-Or, Ariel Shamir, Gal Chechik

Real-Time Neural BRDF with Spherically Distributed Primitives, Yishou Dou, Zhong Zheng, Qiaoqiao Jin, Bingbing Ni, Yugang Chen, Junxiang Ke


Neural Super-Resolution for Real-time Rendering with Radiance Demodulation, Jia Li, Zilong Chen, Xiaolong Wu, Lu Wang, Beibei Wang, Lei Zhang

DiffAvatar: Simulation-Ready Garment Optimization with Differentiable Simulation, Yifei Li, Hsiao-yu Chen, Egor Larionov, Nikolaos Sarafianos, Wojciech Matusik, Tuur Stuyck

Material Palette: Extraction of Materials from a Single Image, Ivan Lopes, Fabio Pizzati, Raoul de Charette

PhysGaussian: Physics-Integrated 3D Gaussians for Generative Dynamics, Tianyi Xie, Zeshun Zong, Yuxing Qiu, Xuan Li, Yutao Feng, Yin Yang, Chenfanyang

Differentiable Point-based Inverse Rendering, Hoon-Gyu Chu, Seokjun Choi, Seung-Hwan Baek


Towards a Perceptual Evaluation Framework for Lighting Estimation, Justine Giroux, Mohammad Reza Karimi Dastjerdi, Yannick Hold-Geoffroy, Javier Vazquez-Corral, Jean-Francois Lalonde

Vector Graphics Generation via Mutually Impulsed Dual-domain Diffusion, Zhongyi Zhao, Ye Chen, Zhangli Hu, Xuanhong Chen, Bingbing Ni

MatFuse: Controllable Material Generation with Diffusion Models, Giuseppe Vecchio, Renato Sortino, Simone Palazzo, Concetto Rampinato

DiffusionLight: Light Probes for Free by Painting a Chrome Ball, Pakkapon Phongthaweew, Waratth Chinchuthakun, Nontaphat Sinsuntithet, Varun Lampani, Amit Raj, Pramook Khungrun, Supasorn Suwajanakorn

PROGRAM GUIDE WEDNESDAY, JUNE 19
417 TexTile: A Differentiable Metric for Texture Tileability, Carlos Rodriguez-Pardo, Dan Casas, Elena Garces, Jorge Lopez-Moreno

418 PIE-NeRF: Physics-based Interactive Elasticodynamics with NeRF, Yutao Feng, Yintong Shang, Xuan Li, Tianjia Shao, Chentianfu Jiang, Yin Yang

419 HashPoint: Accelerated Point Searching and Sampling for Neural Rendering, Jiahao Ma, Miaoanliao Liu, David Ahmed-Aristizabal, Chuong Nguyen

420 3D PaintBrush: Local Stylization of 3D Shapes with Cascaded Score Distillation, Dale Decatur, Itai Langa, Kfir Aberman, Rana Hanocka

421 DUDF: Differentiable Unsigned Distance Fields with Hyperbolic Scaling, Miguel Faiststein, Viviana Slessi, Emmanuel larussi

422 Diffusion 3D Features (Diff3F): Decorating Untextured Shapes with Distilled Semantic Features, Niladri Shekhar Dutt, Sanjeev Muralikrishnan, Niloy J. Mitra

423 LeGO: Leveraging a Surface Deformation Network for Animatable Stylized Face Generation with One Example, Soyeon Yoon, Kwan Yun, Kwangyoon Seo, Sihun Cha, Jang Eun Yoo, Junyong Noh

424 Dr. Bokeh: Differentiable Occlusion-aware Bokeh Rendering, Yichen Sheng, Zixun Yu, Lu Ling, Zhiwen Cao, Xuaner Zhang, Xin Lu, Ke Xian, Haiting Lin, Bedrich Benes

425 DiffInDSceene: Diffusion-based High-Quality 3D Indoor Scene Generation, Xiaoliang Ju, Zhaoyang Huang, Yi Jin, Guofeng Zhang, Yu Qiao, Hongsheng Li

426 LightOctree: Lightweight 3D Spatially-Coherent Indoor Lighting Estimation, Xuecan Wang, Shibang Xiao, Xiaohui Liang

427 SVGDreamer: Text Guided SVG Generation with Diffusion Model, Ximing Xing, Haitao Zhou, Chuang Wang, Jing Zhang, Dong Xu, Qian Yu

428 Control4D: Efficient 4D Portrait Editing with Text, Ruizhi Shao, Jiangxiang Sun, Cheng Peng, Zerong Zheng, Boyao Zhou, Hongwen Zhang, Yebin Liu

429 HumanNorm: Learning Normal Diffusion Model for high- and realistic 3D Human Generation, Xin Huang, Ruizhi Shao, Qi Zhang, Hongwen Zhang, Ying Feng, Yebin Liu, Qing Wang

430 Video2Game: Real-time Interactive Realistic and Browser-Compatible Environment from a Single Video, Hongchi Xia, Zhi-Hao Lin, Wei-Chiu Ma, Shenhong Wang

431 NIVEL: Neural Implicit Vector Layers for Text-to-Vector Generation, Vikas Thamizharasan, Difen Liu, Matthew Fisher, Nanxuan Zhao, Evangelos Kalogerakis, Michal Lukac

432 ESR-NeRF: Emissive Source Reconstruction Using LDR Multi-view Images, Jinseo Jeong, Junseo Koo, Qimeng Zhang, Gunhee Kim

433 DreamPropeller: Supercharge Text-to-3D Generation with Parallel Sampling, Lingzi Zhou, Andy Shih, Chenlin Meng, Stefano Ermon

434 GenesisTex: Adapting Image Denoising Diffusion to Texture Space, Chenjian Gao, Boyan Jiang, Xinghui Li, Yingpeng Zhang, Qian Yu

435 Mosaic-SDF for 3D Generative Models, Lir Yariv, Omri Puny, Ora Ganfi, Yaron Lipman

436 NeRF Analogies: Example-Based Visual Attention Transfer for NeRFs, Michael Fischer, Zhengqian Li, Thu Nguyen-Phuoc, Aljaz Bozic, Zhao Dong, Carl Marshall, Tobias Ritschel

437 Hyper-MD: Mesh Denoising with Customized Parameters Aware of Noise Intensity and Geometric Characteristics, Xingtao Wang, Hongliang Wei, Xiapeng Fan, Debin Zhao

438 QUAdify: Extracting Meshes with Pixel-level Details and Materials from Images, Maximilian Frühaut, Hayko Riemenschneider, Markus Gross, Christopher Schroers

439 SfmCAD: Unsupervised CAD Reconstruction by Learning Sketch-based Feature Modeling Operations, Pu Li, Jianwei Guo, Huibin Li, Bedrich Benes, Dong-Ming Yan

440 Self-Supervised Dual Contouring, Ramana Sundararaman, Roman Klokov, Maks Ovsjanikov

441 SVDTree: Semantic Voxel Diffusion for Single Image Tree Reconstruction, Yutao Liu, Zhihao Liu, Bedrich Benes, Xiaopeng Zhang, Jianwei Guo

442 Text-Conditioned Generative Model of 3D strand-based Human Hairstyles, Vanessa Sklyarova, Egor Zakharov, Otmar Hilliges, Michael J. Black, Justus Thies

443 CAD-SIGNet: CAD Language Inference from Point Clouds using Layer-wise Sketch Instance Guided Attention, Mohammad Sadil Khan, Elona Dupont, Sk Aziz Ali, Kseniya Cherenkova, Ania Kacem, Djamila Rouada

444 Functional Diffusion, Biao Zhang, Peter Wonka

10:30 - 18:45 Art Program (Arch 4CDE)

10:30 - 18:45 DEMOS (Arch 4CDE)

1 DECIMER.ai: An Open-Source Platform for Chemical Structure Recognition from Scientific Literature, Dr. Kohulan Rajan

2 Computer Vision algorithms for 100 skin shades in the beauty industry, Emmanuel Acheampong, Raymond Okeye-Forsorn

3 roboMUA: A predictive and generative AI approach to make up for 100 skin shades, Emmanuel Acheampong


5 RITA: A Real-Time Interactive Talking Avatars Framework, Cheng Wan, Wuxinling Chen, Yupeng Cao

6 ANGEL - Egocentric AR Task Monitoring, Brian Clipp

7 Anyone Can Direct: 3D Film Creation from Text, Zhonggang Cai, Lei Yang, Ziwei Liu

8 Learning Humanoid Locomotion, Ilja Radosavovic

9 WildfireWatch.org: Harnessing AI-Powered Visual Data for Swift and Accurate Wildfire Detection, Kit Merker

10 Enhancing Accuracy and Efficiency of Livestock Monitoring with AI-Powered Visual Data from Plainsight Technologies, Kit Merker

11 Depth Anything, Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashli Feng, Hengshuang Zhao

12 Custom cloth creation and virtual try-on for everyone, Pei Chen, Heng Wang, Zhiyuan Chen, Zhenkun Liu, Shuhua Cao, Li Yang, Minhui Yang

13 Controllable Neural Reconstruction for Autonomous Driving, Péter Kovács, Máté Tóth, Zoltán Bendefy, Zoltán Hortszin, Tamás Matuszka

14 Seeing the World Through Your Eyes, Kevin Zhang, Hadi Alzayer, Sachin Shah

15 Aigen-Al Robotic Solution to Weed Contro, Kayla Stringfellow, Sandra Phillips

16 Clickable Objects - Shoppable Video, Toshiro Ozawa

17 Contactless Optical Vital Sign Monitoring for Elderly Care, Svenja Nicola Kobel, Caroline Ressing, Hueseyin Oenel, Andre Kostfeld, Christian Wiede, Karsten Seidl

11:00 - 13:00 SOCIAL - Challenges/Opportunities for ECRs in Fast Paced AI (Summit Terrace Suite)

**Reservation Required**

12:00 - 14:00 LUNCH Summit ExHall 1-2

13:15 - 14:15 EXPO TRACK - Swami Sivasubramanian Ph.D., Vice President, AI and Data, Amazon Web Services (AWS) (Arch 4F)

13:00 - 14:30 Orals 2A: Image & Video Synthesis (Summit Ballroom)

1 FreeU: Free Lunch in Diffusion U-Net, Chenyang Si, Ziqi Huang, Yuming Jiang, Ziwei Liu

2 Ranni: Taming Text-to-Image Diffusion for Accurate Instruction Following, Yutong Feng, Biao Gong, Di Chen, Yujun Shen, Yu Liu, Jingren Zhou

3 Instruct-Image: Image Generation with Multi-modal Instruction, Hexiang Hu, Kelvin C.K. Chan, Yu-Chuan Su, Wenhui Chen, Yandong Li, Kiyuk Sohn, Yang Zhao, Xue Ben, Boqing Gong, William Cohen, Ming-Wei Chang, Xuhui Jia

4 Attention Calibration for Disentangled Text-to-Image Personalization, Yanbing Zhang, Mengping Yang, Qin Zhou, Zhe Wang

5 Style Aligned Image Generation via Shared Attention, Amir Hertz, Andrey Voynov, Shlomi Fruchter, Daniel Cohen-Or
13:00 - 14:30 Orals 2B: Deep Learning Architectures and Techniques (Summit Flex Hall AB)
1. Neural Redshift: Random Networks are not Random Functions, Damien Teney, Armand Mihai Nicoliciuciu, Valentin Hartmann, Ehsan Abbassnejad
2. Neural Lineage, Runpeng Yu, Xinchao Wang
3. Learning Structure-from-Motion with Graph Attention Networks, Lucas Brynte, José Pedro Iglesias, Carl Olsson, Fredrik Kahl
5. In Search of a Data Transformation That Accelerates Neural Field Training, Junwon Seo, Sangyoon Lee, Kwang In Kim, Jaeho Lee

13:00 - 14:30 Orals 2C: 3D from Multiview and Sensors, (Summit Flex Hall C)
1. Point Transformer V3: Simpler Faster Stronger, Xiaoyang Wu, Li Jiang, Peng-Shuai Wang, Zhijian Liu, Xihui Liu, Yu Qiao, Wanli Ouyang, Tong He, Hengshuang Zhao
2. Matching 2D Images in 3D: Metric Relative Pose from Metric Correspondences, Axel Barroso-Laguna, Sowmya Munukutla, Victor Adrian Prisacariu, Eric Brachmann
3. Seeing the World through Your Eyes, Hadi Alzayer, Kevin Zhang, Brandon Feng, Christopher A. Metzler, Jia-Bin Huang
4. Tri-Perspective View Decomposition for Geometry-Aware Depth Completion, Zhiqiang Yan, Yuankai Lin, Kun Wang, Yupeng Zheng, Yufei Wang, Zhenyu Zhang, Jun Li, Jian Yang
5. Steerers: A Framework for Rotation Equivariant Keypoint Descriptors, Georg Böckman, Johan Edstedt, Michael Felsberg, Fredrik Kahl

14:30 - 15:45 Courtesy Break
15:45 - 16:00 Courtesy Break
16:00 - 17:00 PANEL 1 (Summit Flex Hall ABC)
16:45 - 17:15 Poster Setup Arch 4E
17:00 - 19:00 SOCIAL - How to Balance Research Interests and Academic Tasks (Summit Elliott Bay Room)
**Reservation Required

17:15 - 18:45 Poster Session 2 & Exhibit Hall (Arch 4A-E)
- Highlight paper (check it out)
- Graduate student paper (see graduate sessions)
1. Point Transformer V3: Simpler Faster Stronger, Xiaoyang Wu, Li Jiang, Peng-Shuai Wang, Zhijian Liu, Xihui Liu, Yu Qiao, Wanli Ouyang, Tong He, Hengshuang Zhao
2. Matching 2D Images in 3D: Metric Relative Pose from Metric Correspondences, Axel Barroso-Laguna, Sowmya Munukutla, Victor Adrian Prisacariu, Eric Brachmann
3. Seeing the World through Your Eyes, Hadi Alzayer, Kevin Zhang, Brandon Feng, Christopher A. Metzler, Jia-Bin Huang
4. Tri-Perspective View Decomposition for Geometry-Aware Depth Completion, Zhiqiang Yan, Yuankai Lin, Kun Wang, Yupeng Zheng, Yufei Wang, Zhenyu Zhang, Jun Li, Jian Yang
5. Steerers: A Framework for Rotation Equivariant Keypoint Descriptors, Georg Böckman, Johan Edstedt, Michael Felsberg, Fredrik Kahl

19:00 - 19:30 EfficientDreamer: High-Fidelity and Robust 3D Creation via Orthogonal-view Diffusion Prioros, Zhipeng Hu, Minda Zhao, Chaoyi Zhao, Xinyue Liang, Lincheng Li, Zeng Zhao, Changjie Fan, Xiaowei Zhou, Xin Yu
19:30 - 20:00 Leveraging Camera Triplets for Efficient and Accurate Structure-from-Motion, Lalit Manam, Venu Madhav Govindu
20:00 - 20:30 LAENeRF: Local Appearance Editing for Neuural Radiance Fields, Lukas Radl, Michael Steiner, Andreas Kurz, Markus Steinberger
20:30 - 21:00 SuperPrimitive: Scene Reconstruction at a Primitive Level, Kirill Mazur, Guangbin Bae, Andrew J. Davison
21:00 - 21:30 Revisiting Sampson Approximations for Geometric Estimation Problems, Felix Rydell, Angelica Torres, Viktor Larsson
21:30 - 22:00 Interactive3D: Create What You Want by Interactive 3D Generation, Shaocong Dong, Lihe Ding, Zhanpeng Huang, Zibin Wang, Tianfan Xue, Dan Xu
22:00 - 22:30 Multiples Prior Guided Few-Shot Aerial Scene Rendering, Zihan Gao, Licheng Jiao, Lingling Li, Xu Liu, Fang Liu, Puah Chen, Yuwei Gou
22:30 - 23:00 3DGs-Avatar: Animatable Avatars via deformable 3D Gaussian Splatting, Zhiyin Qin, Shaofei Wang, Marko Mihajlovic, Andreas Geiger, Siyu Tang
23:00 - 23:30 DaReNeRF: Direction-aware Representation for Dynamic Scenes, Ange Lou, Benjamin Planche, Zhengpai Gao, Yamin Li, Tianyu Luan, Hao Ding, Terrence Chen, Jack Noble, Ziyun Wu
23:30 - 24:00 ViewDiff: 3D-Consistent Image Generation with Text-to-Image Models, Lukas Höhlein, Aljaž Božič, Norman Müller, David Novotny, Hung-Tu Tseng, Christian Richardt, Michael Zollhöfer, Matthias Nießner
24:00 - 24:30 LTM: Lightweight Textured Mesh Extraction and Refinement of Large Unbounded Scenes for Efficient Storage and Real-time Rendering, Jaehoon Choi, Rajiv Shah, Qinbo Li, Yipeng Wang, Ayush Saraf, Changil Kim, Jia-Bin Huang, Dinesh Manocha, Suhib Alsisan, Johannes Kopf
24:30 - 25:00 Minimal Perspective Autocorrelation, Andrea Porfiri Dal Cin, Timothy Duff, Luca Magri, Tomas Pajdla
25:00 - 25:30 X-3D: Explicit 3D Structure Modeling for Point Cloud Recognition, Shuofeng Sun, Yongming Rao, Jiwen Lu, Haibin Yan
25:30 - 26:00 2S-UDF: A Novel Two-stage UDF Learning Method for Robust Non-watertight Model Reconstruction from Multi-view Images, Junkai Deng, Fei Hou, Xuhui Chen, Wencheng Wang, Ying He
26:00 - 26:30 UFORec: Generalizable Sparse-View Surface Reconstruction from Arbitrary and Unfavorable Sets, Younju Na, Woo Jae Kim, Kyu Beom Han, Suhyeon Ha, Sung-Eui Yoon
26:30 - 27:00 GenN2N: Generative NeRF2NeRF Translation, Xiangyue Liu, Han Xue, Kunming Luo, Ping Tan, Li Yi
27:00 - 27:30 Text-to-3D Generation with Bidirectional Diffusion using both 2D and 3D priors, Lihe Ding, Shaocong Dong, Zhanpeng Huang, Zibin Wang, Yiyuan Zhang, Kaixiong Gong, Dan Xu, Tianfan Xue
27:30 - 28:00 Noisy One-point Homographies are Surprisingly Good, Yaqing Ding, Jonathan Astemark, Magnus Oskarsson, Viktor Larsson
28:00 - 28:30 Adaptive Multi-Modal Cross-Entropy Loss for Stereo Matching, Peng Xu, Zhiyu Xiang, Chengqiu Yao, Jingyun Fu, Tianyu Pu
28:30 - 29:00 LiDAR4D: Dynamic Neural Fields for Novel Space-time View Synthesis, Zehan Zheng, Fan Lu, Weiyi Xue, Guang Chen, Changjun Jiang
29:00 - 29:30 NC-SDF: Enhancing Indoor Scene Reconstruction Using Neural SDFs with View-Dependent Normal Compensation, Ziyi Chen, Xiaolong Wu, Yu Zhang
29:30 - 30:00 VastGaussian: Vast 3D Gaussians for Large Scene Reconstruction, Jiaqi Lin, Zhihao Li, Xiao Tang, Jianzhuan Liu, Shiyong Liu, Jiaoyu Liu, Yangdi Lu, Xiaofei Wu, Songcen Xu, Youliang Yan, Wenming Yang
30:00 - 30:30 Language-driven Object Fusion into Neural Radiance Fields with Pose-Conditioned Dataset Updates, Ka Chun Shum, Jaeyeon Kim, Binh-Son Hua, Duc Thanh Nguyen, Sai-Kit Yeung
30:30 - 31:00 SPU-PMD: Self-Supervised Point Cloud Upsampling via Progressive Mesh Deformation, Yanzhe Liu, Rong Chen, Yushi Li, Yixi Li, Xuehao Tan
31:00 - 31:30 Intrinsic Image Diffusion for Indoor Single-view Material Estimation, Peter Kocsis, Vincent Sitzmann, Matthias Nießner
Learning Dynamic Tetrahedra for High-Quality Talking Head Synthesis, Zicheng Zhang, Ruobing Zheng, Bonan Li, Congying Han, Tianqi Li, Meng Wang, Tiande Guo, Jingdong Chen, Zewen Liu, Ming Yang

Robust Self-calibration of Focal Lengths from the Fundamental Matrix, Viktor Kocur, Daniel Kyselica, Zuzana Kukelova

RNbNeUS: Reflectance and Normal-based Multi-View 3D Reconstruction, Baptiste Brument, Robin Bruneau, Yvain Quéau, Jean Méloû, François Bernard Lafuze, Jean-Denis Durou, Lilian Calvet

Neural 3D Strokes: Creating Stylized 3D Scenes with Vectorized 3D Strokes, Hao-Bin Duan, Miao Wang, Yan-Xun Li, Yong-Liang Zhang

Unsupervised Template-assisted Point Cloud Shape Correspondence Network, Jiacheng Deng, Jiahao Lu, Tianzhu Zhang

Efficient Detection of Long Consistent Cycles and its Application to Distributed Synchronization, Shaohan Li, Yunpeng Shi, Gilad Lerman

AirPlanes: Accurate Plane Estimation via 3D-Consistent Embeddings, Jamie Watson, Filippo Aleotti, Mohamed Sayed, Zawar Qureshi, Oslan Mac Aodha, Gabriel Brostow, Michael Firman, Sara Vignetti

Accurate Training Data for Occupancy Map Prediction in Automated Driving Using Evidence Theory, Jonas Käible, Sascha Wirges, Maxim Tatarchenko, Eddy Ilg

Continuous Pose for Monocular Cameras in Neural Implicit Representation, Qi Ma, Danda Pani Paudel, Ajad Chhatkuli, Luc Van Gool

Towards 3D Vision with Low-Cost Single-Photon Cameras, Fengzhou Mu, Carter Sifferman, Sasha Jungerman, Yiquan Li, Mark Han, Michael Gleicher, Mohit Gupta, Yin Li

Inlier Confidence Calibration for Point Cloud Registration, Yongzhe Yuan, Yue Wu, Xiaolong Fan, Maqoong Gong, Qiguang Miao, Wenping Ma

GaussianShader: 3D Gaussian Splatting with Shading Functions for Reflective Surfaces, Yingwenqi Jiang, Jiadong Tu, Yuan Liu, Xifeng Gao, Xiaoxiao Long, Wenping Wang, Yuxin Ma

Language Embedded 3D Gaussians for Open-Vocabulary Scene Understanding, Jin-Chuan Shi, Miao Wang, Hao-Bin Duan, Shao-Hua Guan

MVIP-NeRF: Multi-view 3D Inpainting on NeRF Scenes via Diffusion Prior, Honghua Chen, Chen Change Loy, Xingang Pan

SuGaR: Surface-Aligned Gaussian Splatting for Efficient 3D Mesh Reconstruction and High-Quality Mesh Rendering, Antoine Guédon, Vincent Lepetit

DreamControl: Control-Based Text-to-3D Generation with 3D Self-Prior, Gabriel Dogadov, Ugo Finnendahl, Marc Alexa

Feature Interaction for Dense Predictions, Chunlong Xia, Xingjian Wang, Feng Lu, Xin Hao, Yifeng Shi

Lagrangian Particle Optimization, Takahiro Kaneko

DiffusionAvatars: Deferred Diffusion for High-fidelity 3D Head Avatars, Tobias Kirschstein, Simon Glebenhain, Matthias Nießner

VIT-NeMo: Vision Transformer with Convolutional Multi-scale Feature Interaction for Dense Predictions, Yunpeng Shi, Chunlong Xia, Xingjian Wang, Feng Lu, Xin Hao, Yifeng Shi

Pose-Transformed Equivariant Network for 3D Point Trajectory Prediction, Ruixuan Yu, Jian Sun

UniRePLKNet: A Universal Perception Large-Kernel ConvNet for Audio Video Point Cloud Time-Series and Image Recognition, Xiaohan Ding, Yu yuan Zhang, Yixiao Ge, Sijie Zhao, Lin Song, Xiangyu Yue, Ying Shan

KPCovX: Modernizing Kernel Point Convolution with Kernel Attention, Hugues Thomas, Yao-Hung Hubert Tsai, Timothy D. Barfoot, Jian Zhang


Affine Equivariant Networks Based on Different Invariants, Yikang Li, Yeqing Qiu, Yuxuan Chen, Lingshen He, Zhouchen Lin

PeLK: Parameter-efficient Large Kernel ConvNets with Peripheral Convolution, Honghao Chen, Xiangxiang Chu, Yongjian Ren, Xin Zhao, Kaiqi Huang

Making Vision Transformers Truly Shift-Equivariant, Renan A. Rojas-Gomez, Teck-Yian Lim, Minh N. Do, Raymond A. Yeh

Once for Both: Single Stage of Importance and Sparsity Search for Vision Transformer Compression, Hancheng Ye, Chong Yu, Peng Ye, Renju Xia, Yansong Tang, Jiwen Lu, Tao Chen, Bo Zhang

Data-Free Quantization via Pseudo-label Filtering, Chunxiao Fan, Ziqi Wang, Dan Guo, Meng Wang

FedHCA2: Towards Hetero-Client Federated Multi-Task Learning, Yuxiang Lu, Suizhi Huang, Yuwen Yang, Shalayiding Sirejiding, Yue Ding, Hongtao Lu

SpikingRegressor: Bridging ResNet and Vision Transformer in Spiking Neural Networks, Xin Yu, Zecheng Hao, Zhaofei Yu

TetraSphere: A Neural Descriptor for O(3)-Invariant Point Cloud Analysis, Pavlo Melnyk, Andreas Robinson, Michael Felsberg, Mårtén Wadenbäck

Friendly Sharpness-Aware Minimization, Tao Li, Pan Zhou, Zhengbao He, Xinwen Cheng, Xiaolin Huang

RMT: Retentive Networks Meet Vision Transformers, Qihan Fan, Huaibo Huang, Mingrui Chen, Hongmin Liu, Ran He

Efficient Deformable ConvNets: Rethinking Dynamic and Sparse Operator for Vision Applications, Yuwen Xiong, Zhiqi Li, Yuntao Chen, Feng Wang, Xizhou Zhu, Jiapeng Luo, Lewei Lu, Jifeng Dai


Neural Redshift: Random Networks are not Random Functions, Damien Teney, Armand Mihai Nicloșciuciu, Valentijn Hartmann, Ehsan Abbassnejad

InceptionNeXt: When Inception Meets ConvNeXt, Weihao Yu, Pan Zhou, Shuicheng Yan, Xinchao Wang

BiPer: Binary Neural Networks using a Periodic Function, BiPer: Binary Neural Networks using a Periodic Function, BiPer: Binary Neural Networks using a Periodic Function, BiPer: Binary Neural Networks using a Periodic Function, BiPer: Binary Neural Networks using a Periodic Function, BiPer: Binary Neural Networks using a Periodic Function, BiPer: Binary Neural Networks using a Periodic Function, BiPer: Binary Neural Networks using a Periodic Function, BiPer: Binary Neural Networks using a Periodic Function, BiPer: Binary Neural Networks using a Periodic Function, BiPer: Binary Neural Networks using a Periodic Function, BiPer: Binary Neural Networks using a Peri

A&B BNN: Add&Bit-Operation-Only Hardware-Friendly Binary Neural Network, Ruichen Ma, Quanchao Qiao, Yian Liu, Liwei Meng, Ning Ning, Yang Liu, Shaogang Hu

Neural Clustering based Visual Representation Learning, Guikun Chen, Xia Li, Yi Yang, Wenguan Wang

Building Optimal Neural Architectures using Interpretable Knowledge, Keith G. Mills, Fred X. Han, Mohammad Salameh, Shengyao Lu, Chunhua Zhou, Jiao He, Fengyu Sun, Di Niu

Towards More Accurate Diffusion Model Acceleration with A Timestep Tuner, Mengfei Xia, Yujun Shen, Changsong Lei, Yu
Zhou, Deli Zhao, Ran Yi, Wenping Wang, Yong-Jin Liu
99 UniPTVS: A Unified Framework for Proficient Post-Training Sparcity, Jingqing Xie, Yuxin Zhang, Mingbao Lin, Liujian Cao, Rongrong Ji
100 Learning Structure-from-Motion with Graph Attention Networks, Lucas Brynte, José Pedro Iglesias, Carl Olsson, Fredrik Kahl
101 SHVIT: Single-Head Vision Transformer with Memory Efficient Macro Design, Seokju Yun, Youngmin Ro
102 Denoising Point Clouds in Latent Space via Graph Convolution and Invertible Neural Network, Aihua Mao, Biao Yan, Zijing Ma, Ying He
103 JointSQ: Joint Sparserification-Quantization for Distributed Learning, Weiyang Xie, Haowei Li, Jitao Ma, Yunsong Li, Jie Lei, Dongliu Liu, Leyuan Fang
105 RepAn: Enhanced Annealing through Re-parameterization, Xiang Fei, Xiawu Zhang, Yan Wang, Fei Chao, Chenglin Wu, Liujian Cao
106 D^4: Dataset Distillation via Disentangled Diffusion Model, Duo Su, Junjie Hou, Weizhi Gao, Yingjie Tan, Bowen Tang
107 State Space Models for Event Cameras, Nikola Zubic, Mathias Gehrig, Davide Scaramuzza
108 Your Image is My Video: Reshaping the Receptive Field via Image-To-Video Differentiable AutoAugmentation and Fusion, Sofia Casarin, Cynthia I. Ugwu, Sergio Escalera, Oswald Lanz
109 Sparse Semi-DETR: Sparse Learnable Queries for Semi-Supervised Object Detection, Tahira Shehzadi, Khurram Azeem
110 MAPSeg: Unified Unsupervised Domain Adaptation for Heterogeneous Medical Image Classification Based on 3D Masked Autoencoding and Pseudo-Labeling, Xuzhe Zhang, Yuhao Wu, Elsa Angelini, Ang Li, Jia Guo, Jerod M. Rasmussen, Thomas G. O’Connor, Pathik D. Wadhwa, Andrea Porlini, Jackowski, Hai Li, Jonathan Posner, Andrew F. Laine, Yun Wang
111 FedUV: Uniformity and Variance for Heterogeneous Federated Learning, Ha Min Son, Moon-Hyun Kim, Tai-Myoung Chung, Chao Huang, Xin Liu
112 Florence-2: Advancing a Unified Representation for a Variety of 8 Vision Tasks, Bin Xiao, Haiping Wu, Weijian Xu, Xiyang Dai, Houdong Hu, Yumao Lu, Michael Zeng, Ce Liu, Lu Yuan
113 Pick-or-Mix: Dynamic Channel Sampling forConvNets, Ashish Kumar, Daneul Kim, Jaesik Park, Laxmidhar Behera
114 Sheared Backpropagation for Fine-tuning Foundation Models, Zhiyu Yang, Li Sheng, Liang Ding, Xinmei Tian, Yixiang Zhang, Hanpeng Liu, Stephen Lin, Kun He
115 Az-NAS: Assembling Zero-Cost Proxies for Network Architecture Search, Junghyup Lee, Bumsab Ham
116 MRFP: Learning Generalizable Semantic Segmentation from Sim-2-Real with Multi-Resolution Feature Perturbation, Sumanth Udupa, Prajwal Gurunath, Aniruddh Sikdar, Suresh Sundaram
117 Training-Free Pretrained Model Merging, Zhengqi Xu, Ke Yuan, Huiqiong Wang, Yang Wang, Mingli Song, Jie Song
118 Training Generative Image Super-Resolution Models by Wavelet-Domain Losses Enables Better Control of Artifacts, Cansu Korkmaz, A. Murat Tekalp, Zafer Dogan
119 IReNe: Instant Recoloring of Neural Radiance Fields, Alessio Mazzucchelli, Adrian Garcia-Garcia, Elena Garces, Fernando Rivas-Manzaneque, Francesc Moreno-Noguer, Adrian Penate-Sanchez
120 AdaShift: Learning Discriminative Self-Gated Neural Feature Activation With an Adaptive Shift Factor, Sudong Cai
121 Kernel Adaptive Convolution for Scene Text Detection via Distance Map Prediction, Jinzheng Zheng, Heng Fan, Libo Zhang
122 Towards Accurate and Robust Architectures via Neural Architecture Search, Yuwei Ou, Yuqi Feng, Yinan Sun
123 PDF: A Probability-Driven Framework for Open World 3D Point Cloud Semantic Segmentation, Jinfeng Xu, Siyuan Yang, Xianzhi Li, Yuan Tang, Yixue Hao, Long Hu, Min Chen
124 Permutation Equivalence of Transformers and Its Applications, Hengyuan Xu, Liyao Xiang, Hangyu Ye, Dixi Yao, Pengzhi Chu, Baochun Li
125 MedBN: Robust Test-Time Adaptation against Malicious Test Samples, Hyejin Park, Jeongyeon Hwang, Sunung Mun, Sangdon Park, Jangseul Ok
126 Small Scale Data-Free Knowledge Distillation, He Liu, Yikai Wang, Huaping Liu, Fuchun Sun, Anbang Yao
127 Identifying Important Group of Pixels using Interactions, Kosuke Sumiyasu, Kazuhiko Kawamoto, Hiroshi Kera
128 Efficiently Assemble Normalization Layers and Regularization for Federated Domain Generalization, Khiem Le, Long Ho, Cuong Do, Danh Le-Phuc, Koki-Seng Wong
129 OrthCaps: An Orthogonal CapsNet with Sparse Attention Routing and Pruning, Xinyu Geng, Jiawen Wang, Jiawei Gong, Yuerong Xue, Jun Xu, Fanglin Chen, Xiaolin Huang
130 Mean-Shift Feature Transformer, Takumi Kobayashi
131 You Only Need Less Attention at Each Stage in Vision Transformers, Shouxi Zhang, Hanpeng Liu, Stephen Lin, Kun He
132 HEAL-SWIN: A Vision Transformer On The Sphere, Oscar Carlsson, Ian E. Gerken, Hampus Linander, Heiner Speiβ, Fredrik Ohlsson, Christoffer Petersson, Daniel Persson
134 Unlocking the Potential of Prompt-Tuning in Bridging Generalized and Personalized Federated Learning, Wenlong Deng, Christos Thrampoulidis, Xiaoxiao Li
135 MR-VNet: Media Restoration using Volterra Networks, Siddharth Roheda, Amit Unde, Loay Rashid
136 Multimodal Pathway: Improve Transformers with Irrelevant Data from Other Modalities, Yijuan Zhang, Xiaohan Ding, Kaixiong Gong, Yixiao Ge, Ying Shan, Xiangyu Yue
137 GreedyViG: Dynamic Axial Graph Construction for Efficient Vision GNNs, Mustafa Munir, William Avery, Md Mostafijur Rahman, Radu Marculescu
138 FlowerFormer: Empowering Neural Architecture Encoding using a Flow-aware Graph Transformer, Dongyeong Wang, Hyunjoo Kim, Sunwoo Kim, Kijung Shin
139 Mixed-Precision Quantization for Federated Learning on Resource-Constrained Heterogeneous Devices, Huancheng Chen, Haris Vikalo
140 In Search of a Data Transformation That Accelerates Neural Field Training, Junwoo Seo, Sangyoon Lee, Kwang In Kim, Jaeho Lee
141 Wired Perspectives: Multi-View Wire Art Embaces Generative AI, Zhiyu Qu, Lan Yang, Honggang Zhang, Tao Xiang, Kaiyue Pang, Yi-Zhe Song
142 DemoFusion: Democratising High-Resolution Image Generation With No $$$, Ruoyi Du, Dongliang Chang, Timothy Hospedales, Yi-Zhe Song, Zhanyu Ma
144 InteractDiffusion: Interaction Control in Text-to-Image Diffusion Models, Jiun Tian Hoe, Xudong Jiang, Chee Seng Chan, Yap-Peng Tan, Weipeng Hu
145 Intelligent Grimm - Open-ended Visual Storytelling via Latent Diffusion Models, Jiun Tian Hoe, Xudong Jiang, Chee Seng Chan, Yap-Peng Tan, Weipeng Hu
146 ControlledRoom3D: Room Generation using Semantic Proxy Rooms, Jonas Schult, Sam Tsai, Lukas Hölllein, Bichen Wu, Jiailiang Wang, Chih-Yao Ma, Kunpeng Li, Xiaofang Wang, Felix Wimbauer, Zijian He, Peizhao Zhang, Bastian Leibe, Peter Vajda, Ji Hou
147 Cache Me if You Can: Accelerating Diffusion Models through Block Caching, Felix Wimbauer, Bichen Wu, Edgar Schoenfeld, Xiaoliang Dai, Ji Hou, Zijian He, Artsiom Sanakoyeu, Peizhao Zhang, Sam Tsai, Jonas Kohler, Christian Rupprecht, Daniel Cremers, Peter Vajda, Jialiang Wang
139 InstanceDiffusion: Instance-level Control for Image Generation, Xudong Wang, Trevor Darrell, Sai Saketh Rambhatla, Rohit Girdhar, Ishan Misra
140 Make-it-Vivid: Dressing Your Animatable Biped Cartoon Characters from Text, Junshu Tang, Yanhong Zeng, Ke Fan, Xuheng Wang, Bo Dai, Kai Chen, Lizhuang Ma
141 ZONE: Zero-Shot Instruction-Guided Local Editing, Shanglin Li, Bohan Zeng, Yutang Feng, Sicong Gao, Xiuhui Liu, Jianming Liu, Lin Li, Xu Tang, Yao Hu, Jianzhuan Liu, Baochang Zhang
142 Don’t Drop Your Samples! Coherence-Aware Training Benefits Conditional Diffusion, Nicolas Dufour, Victor Besnier, Vicky Kalogeiton, David Picard
143 Generating Illustrated Instructions, Sachit Menon, Ishan Misra, Rohit Girdhar
144 SpikeNeRF: Learning Neural Radiance Fields from Continuous Spike Stream, Lin Zhu, Kangmin Jia, Yifan Zhao, Yunshan Qi, Lizhi Wang, Hua Huang
145 Dancing with Still Images: Video Distillation via Static-Dynamic Disentanglement, Ziyu Wang, Yue Xu, Cewu Lu, Yong-Lu Li
146 UnIGS: Unified Representation for Image Generation and Segmentation, Lu Qi, Lehan Yang, Weidong Guo, Yu Xu, Bo Du, Varun Jampani, Ming-Hsuan Yang
147 Adversarial Text to Continuous Image Generation, Killichbek Haydarov, Aashiq Muhamed, Xiaochen Qian, Jovana Lazarevic, Ivan Skorokhodov, Chaimutdia Jayanga Galappaththige, Mohamed Elhoseiny
148 Self-correcting LLM-controlled Diffusion Models, Tsung-Han Wu, Long Lian, Joseph E. Gonzalez, Boyi Li, Trevor Darrell
149 TINO-Edit: Timestamp and Noise Optimization for Robust Diffusion-Based Image Editing, Sherry X Chen, Yaron Vaxman, Elad Ben Baruch, David Assulin, Aviad Moreshet, Kuo-Chin Lien, Misha Sra, Pradeep Sen
150 Taming Stable Diffusion for Text to 360 Panorama Image Generation, Cheng Zhang, Qinyi Wu, Camilo Cruz Gambardella, Xiaoshui Huang, Dinh Phung, Wanli Ouyang, Jianfei Cai
151 EmoGen: Emotional Image Content Generation with Text-to-Image Diffusion Models, Jingyuan Yang, Jiawei Feng, Hui Huang
152 Carve3D: Improving Multi-view Reconstruction Consistency for Diffusion Models with RL Finetuning, Dasi Xie, Jiahao Li, Hao Tan, Xin Sun, Zhixin Shu, Yi Zhou, Sai Bi, Sören Pirk, Ari E. Kaufman
153 FreeU: Free Lunch in Diffusion U-Net, Chenyang Si, Zhiqiang Huang, Yuming Jiang, Ziwei Liu
154 Move Anything with Layered Scene Diffusion, Jiawei Ren, Mengmeng Xu, Jui-Chieh Wu, Ziwei Liu, Tao Xiang, Antoine Toisoul
155 DiffAgent: Fast and Accurate Text-to-Image API Selection with Robust Diffusion-Based Image Editing, Sherry X Chen, Yaron Vaxman, Elad Ben Baruch, David Assulin, Aviad Moreshet, Kuo-Chin Lien, Misha Sra, Pradeep Sen
157 CCEdit: Creative and Controllable Video Editing via Diffusion Models, Jianmin Bao, Chong Luo, Zhibo Chen, Baining Guo
158 Total Selfie: Generating Full-Body Selfies, Tianwei Yin, Daniel S. Park, Jiajun Wu, Jianmin Bao, Chong Luo, Zhibo Chen, Baining Guo
159 Animate: Zero-shot Object-level Video Customization, Xi Chen, Lianghua Huang, Yu Liu, Yujun Shen, Deli Zhao, Hengshuang Zhao
160 InstanceDiffusion: Instance-level Control for Image Generation, Tsung-Han Wu, Long Lian, Joseph E. Gonzalez, Boyi Li, Trevor Darrell
161 TINO-Edit: Timestamp and Noise Optimization for Robust Diffusion-Based Image Editing, Sherry X Chen, Yaron Vaxman, Elad Ben Baruch, David Assulin, Aviad Moreshet, Kuo-Chin Lien, Misha Sra, Pradeep Sen
162 Taming Stable Diffusion for Text to 360 Panorama Image Generation, Cheng Zhang, Qinyi Wu, Camilo Cruz Gambardella, Xiaoshui Huang, Dinh Phung, Wanli Ouyang, Jianfei Cai
163 EmoGen: Emotional Image Content Generation with Text-to-Image Diffusion Models, Jingyuan Yang, Jiawei Feng, Hui Huang
164 Carve3D: Improving Multi-view Reconstruction Consistency for Diffusion Models with RL Finetuning, Dasi Xie, Jiahao Li, Hao Tan, Xin Sun, Zhixin Shu, Yi Zhou, Sai Bi, Sören Pirk, Ari E. Kaufman
165 SmartMask: Context Aware High-Fidelity Mask Generation for Fine-grained Object Insertion and Layout Control, Jaskirat Singh, Jianming Bao, Chong Luo, Cameron Smith, Zhe Lin, Liang Zheng
166 RAVE: Randomized Noise Shuffling for Fast and Consistent Video Editing with Diffusion Models, Ozgur Kara, Bariscan Kurtkaya, Hirid Yesiltepe, James M. Rehg, Pınar Yanardag
167 LucidDreamer: Towards High-Fidelity Text-to-3D Generation via Interval Score Matching, Yixun Liang, Xing Yang, Jiatao Lin, Haodong Li, Xiaogang Xu, Yingcong Chen
168 HyperDreamBooth: HyperNetworks for Fast Personalization of Text-to-Image Models, Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Wei Wei, Tingbo Hou, Yael Pritch, Neal Wadhwa, Michael Rubinstein, Klir Aberman
169 DreamVideo: Composing Your Dream Videos with Customized Subject and Motion, Yueyi Wei, Shiwei Zhang, Zhiwu Qiang, Hangjie Yuan, Zhiheng Liu, Yu Liu, Yingya Zhang, Jingren Zhou, Hongming Shan
170 SurMo: Surface-based 4D Motion Modeling for Dynamic Human Rendering, Tao Hu, Fangzhong Hou, Ziwei Liu
171 Ranni: Taming Text-to-Image Diffusion for Accurate Instruction Following, Yutong Fung, Biao Gong, Di Chen, Yujun Shen, Yu Liu, Jingren Zhou
172 GenHowTo: Learning to Generate Actions and State Transformations from Instructional Videos, Tomáš Souček, Dima Damen, Michael Wray, Ivan Laptev, Josef Sivic
173 A Recipe for Scaling up Text-to-Video Generation with Text-free Videos, Xiang Wang, Shiwei Zhang, Hangjie Yuan, Zhiwu Qiang, Biao Gong, Yingya Zhang, Yujun Shen, Changxin Gao, Nong Sang
174 WaveFace: Authentic Face Restoration with Efficient Frequency Recovery, Yunqi Miao, Jiankang Deng, Jungong Han
175 AnyDoor: Zero-shot Object-level Image Customization, Xi Chen, Lianghua Huang, Yu Liu, Yujun Shen, Deli Zhao, Hengshuang Zhao
176 ElasticDiffusion: Training-free Arbitrary Size Image Generation through Global-Local Content Separation, Moayed Haji-Alia, Guha Balakrishnan, Vicente Orodzne
177 One-step Diffusion with Distribution-Matching Distillation, Tianwei Yin, Michael Gharbi, Richard Zhang, Eli Shechtman, Frédéric Durand, William T. Freeman, Taesung Park
178 Check Locate Rectify: A Training-Free Layout Calibration System for Text-to-Image Generation, Biao Gong, Siteng Huang, Yutong Fung, Shiwei Zhang, Yuyuan Li, Yu Liu
179 Hierarchical Spatio-temporal Decoupling for Text-to-Video Generation, Zhiwu Qiang, Shiwei Zhang, Jiahuo Wang, Xiang Wang, Yueyi Wei, Yingya Zhang, Changxin Gao, Nong Sang
180 HumanGAN: Text-Driven 3D Human Generation with Gaussian Splatting, Xiao Liu, Xiaohan Zhan, Jiaxiang Tang, Ying Shan, Gang Zeng, Dahuai Lin, Xiaihui Liu, Ziwei Liu
181 WonderJourney: Going from Anywhere to Everywhere, Hong-Xing Yu, Haoyi Duan, Junhua Hur, Kyle Sargent, Michael Rubinstein, William T. Freeman, Forbes Collin, Qing Sun, Noah Snively, Jianwu Wu, Charles Herrmann
183 SIGNeRF: Scene Integrated Generation for Neural Radiance Fields, Jan-Niklas Dillmann, Andreas Engelhardt, Hendrik Lensch
184 VideoBooth: Diffusion-based Video Generation with Image Prompts, Yuming Jiang, Tianxing Wu, Shuai Yang, Chenyang Si, Dahuai Lin, Yu Qiao, Chen Change Loy, Ziwei Liu
185 Total Selfie: Generating Full-Body Selfies, Bowei Chen, Brian Curless, Ira Kemelmacher-Shlizerman, Steven M. Seitz
186 CCEdit: Creative and Controllable Video Editing via Diffusion Models, Ruoyu Feng, Wenming Wang, Yanhui Wang, Yuhui Yuan, Jianmin Bao, Chong Luo, Zhibo Chen, Baining Guo
187 Cinematic Behavior Transfer via NeRF-based Differentiable Filming, Xuekun Jiang, Anyi Rao, Jingbo Wang, Dahuai Lin, Bo Dai
188 Improving Subject-Centric Driver Image Synthesis with Subject-Agnostic Guidance, Kelvin C.K. Chan, Yang Zhao, Xuhui Jia, Ming-Hsuan Yang, Huisheng Wang
189 Drag Your Noise: Interactive Point-based Editing via Diffusion
Semantic Propagation, Hadong Liu, Chenshu Xu, Yifei Yang, Litha Zeng, Shengfeng He

Learning Continuous 3D Words for Text-to-Image Generation, Ta-Ying Cheng, Mathews Gadelha, Thibault Groueix, Matthew Fisher, Radomir Mech, Andrew Markham, Niki Trigoni

CHAIN: Enhancing Generalization in Data-Efficient GANs via lipsCHitz continuity constrained Normalization, Yao Ni, Piotr Koniusz

ViVi-d 1- to- 3: Novel View Synthesis with Video Diffusion Models, Jeong-ki Kwak, Erjun Dong, Yuhe Jin, Hanseok Ko, Shweta Mahajan, Kwang Moo Yi

JeDi: Joint-Image Diffusion Models for Finetuning-Free Personalized Text-to-Image Generation, Yu Zeng, Vishal M. Patel, Haochen Wang, Xun Huang, Ying-Bao Li, Ming-Yu Liu, Yosghal Balaji

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

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

MIGC: Multi-Instance Generation Controller for Text-to-Image

Synthesis, Dewei Zhou, You Li, Fan Ma, Xiaoting Zhang, Yi Yang

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

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

Face2Diffusion for Fast and Editable Face Personalization, Kaede Shiohara, Toshihiko Yamasaki

FreeDrag: Feature Dragging for Reliable Point-based Image Editing, Pengyang Ling, Lin Chen, Pan Zhang, Huaiuan Chen, Yi Jin, Jinjin Zheng

OmniLocalRF: Omnidirectional Local Radiance Fields from Dynamic Videos, Dongyoung Choi, Hyeongjong Jang, Min H. Kim

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

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

Panacea: Panoramic and Controllable Video Generation for Autonomous Driving, Yuqing Wen, Yuncheng Zhao, Yingfei Liu, Fan Jia, Yanhui Wang, Chang Luo, Chi Zhang, Tanciar Wang, Xiaoyan Sun, Xiangyu Zhang

360DGD: Controllable Panorama Video Generation with 360-Degree Video Diffusion Model, Qian Wang, Weiqi Li, Chong Mou, Xinhua Cheng, Jian Zhang

CLIC: Concept Learning in Context, Mehdil Safaei, Arjan Mikaeili, a Patashnik, Daniel Cohen-Or, Ali Mahdavi-Amiri

Z*: Zero-shot Style Transfer via Attention Reweighting, Yingyong Deng, Xiangyu He, Fan Tang, Weiming Dong

Tackling the Singularities at the Endpoints of Time Intervals in Diffusion Models, Pengfei Zhu, Yang Sun, Bing Cao, Qinghua Hu

CosmicMan: A Text-to-Image Foundation Model for Humans, Shikai Li, Janglin Fu, Kaiyuan Liu, Wentao Wang, Kwan-Yee Lin, Wayne Wu

Customize your NeRF: Adaptive Source Driven 3D Scene Editing via Local-Global Iterative Training, Runze He, Shaofei Huang, Xuecheng Nie, Tiairui Hui, Luocz Li, Jiaoi Dai, Jizhong Han, Guanbin Li, Si Liu

PICLE: PhotorealisticIC virtual Try-on from Unconstrained designs, Shuhang Ding, Duomin Wang, Yipeng Qin, Zirong Jin, Baojun Wang, Xiaoguang Han

Focus on Your Instructions: Fine-grained and Multi-instruction Image Editing by Attention Modulation, Qin Guo, Tianwei Lin

Make-Your-Anchor: A Diffusion-based 2D Avatar Generation Framework, Ziyao Huang, Fan Tang, Yong Zhang, Xiaodong Cun, Juan Cao, Jintao Li, Tong-Yee Lee

Revisiting Non-Autoregressive Transformers for Efficient Image Synthesis, Zanlin Ni, Yulin Wang, Renping Zhou, Jiayi Guo, Jinyu Hu, Zhiyuan Liu, Shijie Song, Yuan Yao, Gao Huang

Texture-Preserving Diffusion Models for High-Fidelity Virtual Try-On, Yu Yang, Changdong Ding, Zhihong Xin, Junhuao Huang, Jin Tao, Xianmin Xu

PromptCoT: Align Prompt Distribution via Adapted Chain-of-Thought, Junyi Yao, Yijiang Liu, Zhen Dong, Mingfei Guo, Kelin Hu, Kurt Keutzer, Li Du, Daquan Zhou, Shanghang Zhang

Snap Video: Scaled Spatiotemporal Transformers for Text-to-Video

Synthesis, Willi Menapace, Aliaksandr Siarohin, Ivan Skorokhodov, Ekaterina Deyneka, Tsai-Shen Chen, Anil Kag, Yuwei Fang, Alexei Stoliar, Elisa Ricci, Jian Ren, Sergey Tulyakov


Text-Driven Image Editing via Learnable Regions, Yuanze Lin, Yi-chen Chen, Yi-Hsuan Tsai, Li Jiang, Ming-Hsuan Yang

On Exact Inversion of DPM-Solvers, Seongmin Hong, Kyonghyun Lee, Suh Yoon Jeon, Hyeowon Bae, Se Young Chun

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

ConsistNet: Enforcing 3D Consistency for Multi-view Images Diffusion, Jiayu Yang, Zichang Cheng, Yunfei Duan, Pan Ji, Hongdong Li

LAMP: Learn A Motion Pattern for Few-Shot Video Generation, Ruiqi Wu, Liangyu Chen, Tong Yang, Chunle Guo, Chongyi Li, Xiangyu Zhang

Task-Customized Mixture of Adapters for General Image Fusion, Pengfei Zhu, Yang Sun, Bing Cao, Qinghua Hu

Beyond Textual Constraints: Learning Novel Diffusion Conditions with Fewer Examples, Xuyang Yu, Bingzhen Liu, Chenxi Zheng, Kun Quan, Zhiyuan Xu, Haowei Jiang, Shengfeng He

Portrait4D: Learning One-Shot 4D Head Avatar Synthesis using Synthetic Data, Yu Deng, Duomin Wang, Xiaohang Ren, Xingyu Chen, Baoyuan Wang

Animating General Image with Large Visual Motion Model, Dengsheng Chen, Xiaojun Wei, Xiaolin Wei

Sat2Scene: 3D Urban Scene Generation from Satellite Images

with Diffusion, Zuoyue Li, Zhenqiang Li, Zhaoqin Cui, Marc Pollefeys, Martin R. Oswald

Seeing and Hearing: Open-domain Visual-Audio Generation with Diffusion Latent Aligners, Yazhou Xing, Yingqing He, Zeyue Tian, Xintao Wang, Qifeng Chen

AVID: Any-Length Video Inpainting with Diffusion Model, Zhixing Zhang, Bichen Wu, Xiaoyan Wang, Yaqiao Luo, Luxin Zhang, Yiran Zhao, Peter Vajda, Dimitris Metaxas, Licheng Yu

Generative Powers of Ten, Xiaojuan Wang, Jannie Konkanen, Brian Curless, Steven M. Seitz, Ira Kemelmacher-Shlizerman, Ben Mildenhall, Pratul Srinivasan, Dor Verbin, Aleksander Holynski

DistriFusion: Distributed Parallel Inference for High-Resolution Diffusion Models, Muyang Li, Tianle Cai, Jiaxin Cao, Qinsheng Zhang, Han Cai, Junjie Bai, Yangqing Jia, Kai Li, Song Han

Condition-Aware Neural Network for Controlled Image Generation, Han Cai, Muyang Li, Qinsheng Zhang, Ming-Yu Liu, Song Han

It’s All About Your Sketch: Democratising Sketch Control in Diffusion Models, Subhadeep Koley, Ayan Kumar Bhunia, Deepanshu Sekhri, Aneesah Sain, Pinaki Nath Chowdhury, Tao Xiang, Yi-Zhe Song

FaceChain-SuDe: Building Derived Class to Inherit Category Attributes for One-shot Subject-Driven Generation, Pengchong Qiao, Lei Shang, Chang Liu, Baigui Sun, Xiangyang Ji, Jie Chen

PanNet-In-Out: Faithful 3D GAN Inversion with Volumetric Decomposition for Face Editing, Yiran Xu, Zhiqin Shu, Cameron Smith, Seoug Wug Oh, Jia-Bin Huang

Video Prediction by Modeling Videos as Continuous Multi-Dimensional Processes, Gaurav Shrivastava, Abhinav Shrivastava

239 Structure-Guided Adversarial Training of Diffusion Models, Ling Yang, Haotian Qian, Zhisong Zhang, Jingwei Liu, Bin Cui

240 Learning Adaptive Spatial Coherent Correlations for Speech- Preserving Facial Expression Manipulation, Tianshui Chen, Jianman Lin, Zhiqing Yang, Chunmei Qing, Liang Lin

241 On the Content Bias in Fréchet Video Distance, Songwei Ge, Aniruddha Mahapatra, Gaurav Parmar, Jun-Yan Zhu, Jia-Bin Huang

242 Residual Learning in Diffusion Models, Junyu Zhang, Daochang Liu,

243 A Unified Approach for Text- and Image-guided 4D Scene Generation, Yufeng Zheng, Xueting Li, Koki Nagano, Sifei Liu, Otmar Hilliges, Shalini De Mello

244 VideoCrafter2: Overcoming Data Limitations for High-Quality Video Diffusion Models, Haoxin Chen, Yong Zhang, Xiaodong Cun, Menghan Xia, Xintao Wang, Chao Weng, Ying Shan

245 Neural Implicit Morphing of Face Images, Guilherme Schardong, Tiago Novello, Hallison Paz, Iurii Medvedev, Vinicius da Silva, Luiz Velho, Nuno Gonçalves

246 One More Step: A Versatile Plug-and-Play Module for Rectifying Diffusion Schedule Flaws and Enhancing Low-Frequency Controls, Minghui Hu, Jianbin Zheng, Chunxia Zhang, Chaoyue Wang, Dacheng Tao, Tat-Jen Cham

247 Video Interpolation with Diffusion Models, Siddhant Jain, Daniel Watson, Eric Tabellion, Aleksander Holsny, Ben Poole, Janne Kontkanen

248 DiffSHEG: A Diffusion-Based Approach for Real-Time Speech-driven Holistic 3D Expression and Gesture Generation, Junhong Chen, Yunfei Liu, Jianan Wan, Ailing Zeng, Yu Li, Qifeng Chen

249 TFMQ-DM: Temporal Feature Maintenance Quantification for Diffusion Models, Yushi Huang, Ruihao Gong, Jing Liu, Tianlong Chen, Xianglong Liu

250 Improving Training Efficiency of Diffusion Models via Multi-Stage Framework and Tailored Multi-Decoder Architecture, Huijie Zhang, Yifu Lu, Ismail Alkhoury, Saiprasad Ravishankar, Dogyoon Song, Qing Qu

251 Scaling Laws of Synthetic Images for Model Training ... for Now, Lijie Fan, Kaifeng Chen, Dilip Krishnan, Dina Katabi, Phillip Isola, Yonglong Tian


253 MaskINT: Video editing via Interpolative Non-autoregressive Masked Transformers, Haoyu Ma, Shahnaz Mahdizadehaghdam, Bichen Wu, Zhipeng Fan, Yuchao Gu, Wenliang Zhao, Lior Shapira, Xiaohui Xie

254 Pose-Adapted Shape Learning for Large-Pose Face Reenactment, Gee-Sern Jison Hsu, Jie-Ying Zhang, Huang Yu Hsiang, Wei-Jie Hong

255 PRDP: Proximal Reward Difference Prediction for Large-Scale Reward Finetuning of Diffusion Models, Fei Deng, Qiwei Wang, Wei Wei, Tingbo Hou, Matthias Grundmann

256 Discriminative Probing and Tuning for Text-into-Image Generation, Leigang Qu, Wenjie Wang, Yongqi Li, Hanwang Zhang, Liqiang Nie, Tat-Seng Chua

257 Towards Automated Movie Trailer Generation, Davit Mureja, Jussi Keppo, Ying Shan, Mike Zheng Shou

258 Dysen-VDM: Empowering Dynamics-aware Text-to-Video Diffusion with LLMs, Hao Fei, Shengqiong Wu, Wei Ji, Hanwang Zhang, Tat-Seng Chua

259 Geometry-aware Reconstruction and Fusion-refined Rendering for Generalizable Neural Radiance Fields, Tianqi Liu, Xinyi Ye, Min Shi, Zihao Huang, Zhiyu Pan, Zhan Peng, Zhiguo Cao

260 DynVideo-E: Harnessing Dynamic NeRF for Large-Scale Motion- and View-Change Human-Centric Video Editing, Jia-Wei Liu, Yan-Pei Cao, Jun Zhang, Qing Zhu, Rui Zhao, Yibin Yin, Jia Wei, Ying Shao, Mike Zheng Shou

261 High-fidelity Person-centric Subject-to-Image Synthesis, Yibin Wang, Weizhong Zhang, Jianwei Zheng, Cheng Jin

262 Relation Recfstion in Diffusion Model, Yinwei Xu, Xinyi Yang, Xinchao Wang

263 Diffusion Handles Enabling 3D Edits for Diffusion Models by Lifting Activations to 3D, Karran Pandey, Paul Guerrero, Mathew Gadelha, Yannick Hold-Geoffroy, Karan Singh, Niyol J. Mitra

264 LeftRefill: Filling Right Canvas based on Left Reference through Generalized Text-to-Image Diffusion Model, Chenjie Cao, Yunuo Cai, Qirole Dong, Yikai Wang, Yanwei Fu

265 FSRT: Facial Scene Representation Transformer for Face Reenactment from Factorized Appearance Head-pose and Facial Expression Features, Andre Rochow, Max Schwarz, Sven Behnke

266 Tailored Visions: Enhancing Text-to-Image Generation with Personalized Prompt Rewriting, Zijie Chen, Lichao Zhang, Fangsheng Weng, Lili Pan, Zhenzhong Lan

267 MMA-Diffusion: MultiModal Attack on Diffusion Models, Yijun
Yang, Ruiyuan Gao, Xiaosen Wang, Tsung-Yi Ho, Nan Xu, Qiang Xu

PIA: Your Personalized Image Animator via Plug-and-Play Modules in Text-to-Image Models, Yiming Zhang, Zhengxing Xing, Yangxiaogang Fang, Kai Chen

Codebook Transfer with Part-of-Speech for Vector-Quantized Image Modeling, Baoquan Zhang, Huabin Wang, Chuyao Luo, Xutao Li, Quotao Liang, Yuming Ye, Xiaochen Qi, Yao He

Generating Non-Stationary Textures using Self-Rectification, Yang Zhou, Rongqun Xiao, Dani Lischinski, Daniel Cohen-Or, Hui Huang

Fast ODE-based Sampling for Diffusion Models in Around 5 Steps, Zhenyu Zhou, Defang Chen, Can Wang, Chun Chen

Deformable One-shot Face Stylization via DINO Semantic Guidance, Yang Zhou, Zichong Chen, Hui Huang

Learning Disentangled Identifiers for Action-Customized Text-to-Image Generation, Siteng Huang, Biao Gong, Yutong Feng, Xi Chen, Yuqian Fu, Yu Liu, Donglin Wang

SwiftBrush: One-Step Text-to-Image Diffusion Model with Variational Score Distillation, Thuan Hoang Nguyen, Anh Tran

Towards Understanding Cross and Self-Attention in Stable Diffusion for Text-Guided Image Editing, Bingyan Liu, Chengyu Wang, Tingfeng Cao, Kui Jia, Jun Huang

SimDA: Simple Diffusion Adapter for Efficient Video Generation, Zhen Xing, Qi Dai, Han Hu, Zuxuan Wu, Yu-Gang Jiang

Unlocking Pre-trained Image Backbones for Semantic Image Synthesis, Tariq Berrada Ifriqi, Jakob Verbeek, Camille Couprie, Karthik Alahari

Shadow-Enlightened Image Outpainting, Hang Yu, Ruilin Li, Shaorong Xie, Jiaoyan Qiu


StyleCineGAN: Landscape Cinematography Generation using a Pre-trained StyleGAN, Jongwoo Choi, Kwanggyoon Seo, Amir Samani Ashari, Junyoung Noh

MotionEditor: Editing Video Motion via Content-Aware Diffusion, Shuyuan Tu, Qi Dai, Zhi-Qi Cheng, Han Hu, Xintong Han, Zuxuan Wu, Yu-Gang Jiang

DanceCamera3D: 3D Camera Movement Synthesis with Music and Dance, Zixuan Wang, Jia Jia, Shikun Sun, Haozhe Wu, Rong Han, Zhenyu Li, Di Tang, Jiaqing Zhou, Jiebo Luo

Diversity-aware Channel Pruning for StyleGAN Compression, Jiwoon Chung, Sangheek Hyun, Sang-Heon Shim, Jae-Pil Heo

DiffMorpher: Unleashing the Capability of Diffusion Models for Image Morphing, Kaiwen Zhang, Yifan Zhou, Xudong Xu, Bo Dai, Xingang Pan

StegoGAN: Leveraging Steganography for Non-Bijective Image-to-Image Translation, Sidr Wu, Yizi Chen, Samuel Mermet, Lorenz Hurni, Konrad Schindler, Nicolas Gonthier, Loic Landrieu

Grounded Text-to-Image Synthesis with Attention Refocusing, Quynh Phung, Songwei Ge, Jia-Bin Huang

VecFusion: Vector Font Generation with Diffusion, Vikas Thamizharasan, Difan Liu, Shantanu Agarwal, Zhenyu Zhou, Defang Chen, Can Wang, Chun Chen

PROGRAM GUIDE

Ryan Po, Guandao Yang, Kfir Aberman, Gordon Wetzstein

Vikas Thamizharasan, Difan Liu, Shantanu Agarwal, Zhenyu Zhou, Defang Chen, Can Wang, Chun Chen

Tulyakov, Jeong Joon Park, Andrea Tagliasacchi, David B. Lindell

❉ Yangbo Chen, Oliver Wang, Richard Zhang, Eli Shechtman, Xiaolong Wang, Michael Gharbi

Learning Multi-Dimensional Human Preference for Text-to-Image Generation, Siyao Zhang, Bohan Wang, Junqiang Wu, Yan Li, Tingting Gao, Di Zhang, Zhongyuan Wang

Dynamic Policy-Driven Adaptive Multi-Instance Learning for Whole Slide Image Classification, Tingting Zhang, Kui Jiang, Hongxuan Yao

Structure Matters: Tackling the Semantic Discrepancy in Diffusion Models for Image Inpainting, Haipeng Liu, Yang Wang, Biao Qian, Meng Wang, Yong Rui

IMPRIENT: Generative Object Compositing by Learning Identity-Preserving Representation, Yizhi Song, ZhiFei Zhang, Zhe Lin, Scott Cohen, Brian Price, Jianming Zhang, Soo Ye Kim, He Zhang, Wei Xiong, Daniel Aligia

Puff-Net: Efficient Style Transfer with Pure Content and Style Feature Fusion Network, Szhe Zhang, Pan Gao, Peng Zhou, Jie Qin

SSR-Encoder: Encoding Selective Subject Representation for Subject-Driven Generation, Yuxuan Zhang, Yiren Song, Jianming Liu, Rui Wang, Jindi Peng, Yu Hao, Huaxia Li, Xu Tang, Yau Ho, Han Pan, Zhongliang Jing

PEEKABOO: Interactive Video Generation via Masked-Diffusion, Yash Jain, Anshul Nasery, Vibhav Vineet, Harkirat Behl

CoDeF: Content Deformation Fields for Temporally Consistent Video Processing, Hao Ouyang, Qiuyu Wang, Yuxi Xiao, Qingyuan Bai, Juntao Zhang, Kecheng Zhang, Xiaowei Zhou, QiFeng Chen, Yujun Shen

DreamMatcher: Appearance Matching Self-Attention for Semantically-Consistent Text-to-Image Personalization, Jisu Nam, Heesu Kim, Dongjae Lee, Siyon Jin, Seunggyong Kim, Seunggyu Chang

DreamComposer: Controllable 3D Object Generation via Multi-View Conditions, Yunhan Yang, Yukun Huang, Xiaoyang Wu, Yuan-Chen Guo, Song-Hai Zhang, Hengshuang Zhao, Tong He, Kihai Liu

Shadow Generation for Composite Image Using Diffusion Model, Qingyang Liu, Junqi You, Jianming Wang, Xin Hao, Bo Zhang, Li Niu

Adversarial Score Distillation: When score distillation meets GAN, Min Wei, Jingkai Zhou, Junyao Sun, Xuesong Zhang

Uncertainty-Aware Source-Free Adaptive Image Super-Resolution with Wavelet Augmentation Transformer, Yuangui Zou, Haoibo Huang, Lei Zhang, Ran He

Animate Anyone: Consistent and Controllable Image-to-Video Synthesis for Character Animation, Li Hu

Person in Place: Generating Associative Skeleton-Guidance Maps for Human-Object Interaction Image Editing, ChangHee Yang, ChanHee Kang, Kyeongbo Kang, Hanni Oh, Suk-Ju Kang

StableVITON: Learning Semantic Correspondence with Latent Diffusion Model for Virtual Try-On, Jeongho Kim, Guojung Gu, Minho Park, Sunghyun Park, Jaegul Choo

Attention Calibration for Disentangled Text-to-Image Personalization, Yanbing Zhang, Mengping Yang, Qin Zhou, Zhe Wang

Personalized Residuals for Concept-Driven Text-to-Image Generation, Cusush Ham, Matthew Fisher, James Hays, Nicholas Kolkin, Yuchen Liu, Richard Zhang, Tobias Hinz

UFOGen: You Forward Once Large Scale Text-to-Image Generation via Diffusion GANs, Yannwu Xu, Yang Zhao, Zhisheng Xiao, Tingbo Hou

FlowVis: Taming Imperfect Optical Flows for Consistent Video-to-

Video Synthesis, Feng Liang, Bichen Wu, Jialiang Wang, Licheng Yu, Kunpeng Li, Yinan Zhao, Isan Misra, Jia-Bin Huang, Peizhao Zhang, Peter Vajda, Diana Marculescu

Readout Guidance: Learning Control from Diffusion Features, Grace Luo, Trevor Darrell, Oliver Wang, Dan B Goldman, Aleksander Holynski

Diffusion Model Alignment Using Direct Preference Optimization, Bram Wallace, Meihua Dang, Rafael Rafailov, Linqi Zhou, Aaron Lou, Senthil Purushwalkam, Stefano Ermon, Caiming Xiong, Shaqiq Joty, Nikhil Naik

Diffusion Models Without Attention, Jing Nathan Yan, Jiatao Gu, Alexander M. Rush
335 CommonCanvas: Open Diffusion Models Trained on Creative-Commons Images, Aaron Gokaslan, A. Feder Cooper, Jasmine Collins, Landan Seguin, Austin Jacobson, Mihir Patel, Jonathan Frankie, Cory Stephenson, Volodymyr Kuleshov


337 Edit One for All: Interactive Batch Image Editing, Thao Nguyen, Utkarsh Ojha, Yuheng Li, Haotian Liu, Yong Jae Lee

338 Wavelet-based Fourier Information Interaction with Frequency Diffusion Adjustment for Underwater Image Restoration, Chen Zhao, Weiling Cai, Chenyu Dong, Chengwei Hu

339 Accelerating Diffusion Sampling with Optimized Time Steps, Shuchen Xue, Zhaohuang Liu, Fei Chen, Shifeng Zhang, Tianyang Hu, Enze Xie, Zhenguo Li


341 Selectively Informative Description can Reduce Undesired Embellishment Entanglements in Text-to-Image Personalization, Jiyeong Kim, Junwong Park, Wanjong Roe

342 Observation-Guided Diffusion Probabilistic Models, Junhong Kang, Jinyoung Choi, Sungik Choi, Bohyung Han

343 Scaling Up Video Summarization Pretraining with Large Language Models, Dawit Mureja Argaw, Seunghyun Yoon, Fabian Caba Heilbron, Hanieh Deilamsalehy, Trung Bui, Zhaowen Wang, Franck Dernoncourt, Joon Son Chung


345 Clockwork Diffusion: Efficient Generation With Model-Step Distillation, Amirkhossein Habibian, Amir Ghodrati, Noor Fathima, Guillaume Sautiere, Rishesh Garrepalli, Fathil Porioki, Jens Petersen

346 SmartEdit: Exploring Complex Instruction-based Image Editing with Multimodal Large Language Models, Yuzhou Huang, Liangbin Xie, Xintao Wang, Ziyang Yuan, Rui Zhu, Yingwei Pan, Yehao Li, Ting Yao, Zhiyuan Zhao, Weiling Cai, Chenyu Dong, Chengwei Hu

347 CAT-DM: Controllable Accelerated Virtual Try-on with Diffusion Model, Jianhao Zeng, Dan Song, Weizhi Nie, Hongshuo Tian, Tongtong Wang, An-An Liu

348 Exact Fusion via Feature Distribution Matching for Few-shot Image Generation, Yingbo Zhou, Yutong Ye, Pengyu Zhang, Xianwei Wei, Mingsong Chen

349 Clockwork Initialization for Face Personalization of Text-to-Image Models, Lianyu Pang, Jinjun Yu, Huanyoung Kim, Chong Mou, Xintao Wang, Ziyang Yuan, Rui Zhu, Yingwei Pan, Yehao Li, Ting Yao, Zhiyuan Zhao, Weiling Cai, Chenyu Dong, Chengwei Hu

350 EasyDiff: Efficient Point-based Manipulation on Diffusion Models, Zhongwei Zhang, Yinghao Shen, Shaoteng Li, Aniruddha Srikar, Hugh Chen, Yifan Zhou, Ziwei Liu, Chen Change Loy, Yuchen Zhang, Yao Yao

351 Exact Fusion via Feature Distribution Matching for Few-shot Image Generation, Yingbo Zhou, Yutong Ye, Pengyu Zhang, Xianwei Wei, Mingsong Chen

352 Cross Initialization for Face Personalization of Text-to-Image Models, Lianyu Pang, Jinjun Yu, Huanyoung Kim, Chong Mou, Xintao Wang, Ziyang Yuan, Rui Zhu, Yingwei Pan, Yehao Li, Ting Yao, Zhiyuan Zhao, Weiling Cai, Chenyu Dong, Chengwei Hu

353 EasyDiff: Efficient Point-based Manipulation on Diffusion Models, Zhongwei Zhang, Yinghao Shen, Shaoteng Li, Aniruddha Srikar, Hugh Chen, Yifan Zhou, Ziwei Liu, Chen Change Loy, Yuchen Zhang, Yao Yao

354 SD-DIT: Unleashing the Power of Self-supervised Discrimination in Diffusion Transformer, Rui Zhu, Yingwei Pan, Yehao Li, Ting Yao, Zhelong Sun, Tao Mei, Chang Wen Chen

355 Towards Effective Usage of Human-Centric Priors in Diffusion Models for Text-based Human Image Generation, Junyang Wang, Zhenhong Sun, Zhiyu Tan, Xuanbai Chen, Weihua Chen, Hao Li, Cheng Zhang, Yang Song

356 Text2QR: Harmonizing Aesthetic Customization and Scanning Robustness for Text-Guided QR Code Generation, Guangyang Wu, Xiaohong Liu, Jun Jia, Xuehua Cui, Guangtuo Zhai

357 Space-Time Diffusion Features for Zero-Shot Text-Driven Motion Transfer, Danah Yatzin, Rafail Fridman, Omer Bar-Tal, Yoni Kasten, Tali Dekel

358 Video Frame Interpolation via Direct Synthesis with the Event-based Reference, Yuhuan Liu, Yajian Deng, Hao Chen, Zhen Yang

359 DiffEditor: Boosting Accuracy and Flexibility on Diffusion-based Image Editing, Chong Mou, Xintao Wang, Jiechong Song, Ying Shan, Jian Zhang

360 EMOPortraits: Emotion-enhanced Multimodal One-shot Head Avatars, Nikita Drobyshev, Antoni Bigata Casademunt, Konstantinos Vougioukas, Zoe Landgraf, Stavros Petridis, Majda Pantic

361 Spacetime Gaussian Feature Splitting for Real-Time Dynamic View Synthesis, Zhan Li, Zhang Chen, Zhong Li, Yi Xu

362 HOIDiffusion: Generating Realistic 3D Hand-Object Interaction Data, Mengzi Zhang, Yang Fu, Zheng Ding, Sifei Liu, Zhourwen Tu, Xiaolong Wang

362 Learned Representation-Guided Diffusion Models for Large-Image Generation, Alexandros Graikos, Srikar Yellapragada, Minhan Le, Saarthak Kapse, Prateek Prasanna, Joel Saltz, Dimitris Samaras

363 InstantBoost: Personalized Text-to-Image Generation without Test-Time Finetuning, Jing Shi, Wei Xiong, Zhe Lin, Hyun Joon Jung

363 TokenCompose: Text-to-Image Diffusion with Token-level Supervision, Zirui Wang, Zhihouda Sha, Zheng Ding, Yilin Wang, Zhourwen Tu

364 Geometry Transfer for Styling Radiation Fields, Hyunyoung Jung, Seonghyeon Nam, Nikolaos Sarafianos, Sungjoo Yoo, Alexander Sorkine-Hornung, Rakesh Ranjan

365 Align Your Gallant: Text-to-4D with Dynamic 3D Gallants and Composed Diffusion Models, Huan Ling, Seung Wook Kim, Antonio Torralba, Sanja Fidler, Karsten Kreis

366 DreamSalon: A Staged Diffusion Framework for Preserving Identity-Context in Editable Face Generation, Haonan Lin

367 Video-P2P: Video Editing with Cross-attention Control, Shoahtong Liu, Yuechen Zhang, Wenbo Li, Zhe Lin, Jiaya Jia

368 PAIR Diffusion: A Comprehensive Multimodal Object-Level Image Editor, Vidit Goel, Elia Peruzzo, Yifan Jiang, Deja Xu, Xinqian Xu, Nicu Sebe, Trevor Darrell, Zhangyang Wang, Humphrey Shi

369 ArtAdapter: Text-to-Image Style Transfer using Multi-Level Style Encoder and Explicit Adaptation, Dar-Yen Chen, Hamish Tennent, Ching-Wen Hsu

370 DemoCaricature: Democarting Caricature Generation with a Rough Sketch, Dar-Yen Chen, Ayan Kumar Bhunia, Subhadeep Koley, Aneesha Sain, Pinaki Nath Chowdhury, Yi-Zhe Song

371 PhotoMaker: Customizing Realistic Human Photos via Stackeds 3D Embedding, Zhen Li, Mindeng Cao, Xintao Wang, Zhongqiang Qi, Ming-Ming Cheng, Ying Shan

372 Predicated Diffusion: Predicate Logic-Based Attention Guidance for Text-to-Image Diffusion Models, Kotsa Sueyoshi, Takashi Matsubara

373 SNED: Superposition Network Architecture Search for Efficient Video Diffusion Model, Zhengliang Li, Yan Kang, Yuchen Liu, Difan Liu, Tobias Hinz, Feng Liu, Yanzi Wang

374 TRIP: Temporal Residual Learning with Image Noise Prior for Image-to-Image Diffusion Models, Zhongwei Zhang, Fucheng Long, Yingwei Pan, Zhaofan Qiu, Ting Yao, Yang Cao, Tao Mei

375 Prompt-Free Diffusion: Taking "Text" out of Text-to-Image Diffusion Models, Xingqian Xu, Jiayi Guo, Zhangyang Wang, Gao Huang, Irfan Essa, Humphrey Shi

376 DEADiff: An Efficient Stylization Diffusion Model with Disentangled Representations, Tianhao Q1, Shancheng Fang, Yanze Wu, Hongtao Xie, Jiawei Liu, Long Chen, Qian He, Yongdong Zhang

377 FRESCO: Spatial-Temporal Correspondence for Zero-Shot Video Translation, Shuai Yang, Yifan Zhou, Ziwei Liu, Chen Change Loy

378 Correcting Diffusion Generation through Resampling, Yujian Liu, Yang Zhang, Tommi Jaakkola, Shiyu Chang

379 AnyScene: Customized Image Synthesis with Composited Foreground, Ruidong Chen, Lanjun Wang, Weizhi Nie, Yongdong Zhang, An-An Liu

380 Grid Diffusion Models for Text-to-Video Generation, Taegyeong Lee, Soyeong Kwon, Taehwan Kim

381 Direct2.5: Diverse Text-to-3D Generation via Multi-view 2.5D Diffusion, Yuanxun Lu, Jingyang Zhang, Shiewei Li, Tian Fang, David McKinnon, Yanghai Tsin, Long Quan, Xun Cao, Yao Yao

382 Anomaly Score: Evaluating Generative Models and Individual Generated Images based on Complexity and Vulnerability, Jaehui Hwang, Junghyuk Lee, Jong-Seok Lee
PROGRAM GUIDE

WEDNESDAY, JUNE 19

384 Style Aligned Image Generation via Shared Attention,
*Amir Hertz, Andrey Vovnyov, Shlomi Fruchter, Daniel Cohen-Or

385 Zero-Painter: Training-Free Layout Control for Text-to-Image Synthesis, 
Mariansa Ohanian, Hayk Manukyan, Zhangyang Wang, Shant Navasaryan, Humphrey Shi

386 X-Adapter: Adding Universal Compatibility of Plugins for Upgraded 
Diffusion Model, Lingmin Ran, Xiaodong Cun, Jia-Wei Liu, Rui Zhao, 
Song Zijie, Xintao Wang, Jussi Keppo, Mike Zheng Shou

387 Neural Point Cloud Diffusion for Disentangled 3D Shape and 
Appearance Generation, Philipp Schröppel, Christopher Wewer, 
Jan Eric Lenssen, Eddy Igl, Thomas Brox

388 Style Injection in Diffusion: A Training-free Approach for Adapting
Large-scale Diffusion Models for Style Transfer, Jiwoo Chung, 
Sangeek Hyun, Jae-Pil Heo

389 Vlogger: Make Your Dream A Vlog, Shaobin Zhuang, Kunchang Li, 
Xinyin Chen, Yaohui Wang, Ziwei Liu, Yu Qiao, Yali Wang

390 Faces that Speak: Jointly Synthesizing Talking Face and Speech 
from Text, Youngjoon Jang, Ji-Hoon Kim, Junseok Ahn, Doyeop Kwak, 
Hong-Sun Yang, Yoon-Chool Ji, Il-Hwan Kim, Byeong-Yeo Kim, 
Joon Son Chung

391 Prompt Augment for Self-supervised Text-guided Image 
Manipulation, Rumeysa Bodur, Binod Bhattarai, Tae-Kyun Kim

392 DragDiffusion: Harnessing Diffusion Models for Interactive Point-
based Image Editing, Yujin Shi, Chuhui Xue, Jun Hao Liew, Jiachun Pan, 
Hanshu Yan, Wenqing Zhang, Vincent Y. F. Tan, Song Bai

393 Make Pixels Dance: High-Dynamic Video Generation, Yan Zeng, 
Guoqiang Wei, Jiliang Zhang, Jin Xiong, Zechen Zhang, Hang Li

394 LEDITS++: Limitless Image Editing using Text-to-Image 
Models, Manuel Brack, Felix Friedrich, Katharina Kornmeier, Linoy Tsaban, Patrick Schramowski, Kristian Kersting, 
Apolinar Passos

395 Emu Edit: Precise Image Editing via Recognition and Generation
Tasks, Shelly Sheynin, Adam Polyak, Uziel Singer, Yuval Kirshtain, 
Amir Zilber, Oron Ashual, Devi Parikh, Yaniv Taigman

396 Concept Weaver: Enabling Multi-Concept Fusion in Text-to-Image 
Models, Gihyun Kwon, Simon Jenni, Dingzeyu Li, Joon-Young Lee, 
Jong Chul Ye, Fabian Caba Heilbron

397 ACT-Diffusion: Efficient Adversarial Consistency Training for One-
step Diffusion Models, Fei Kong, Jinhuan Duan, Lichao Sun, Hao Cheng, 
Renjing Xu, Hengtao Shen, Xiaofeng Zhu, Xiaoshuang Shi, 
Kaidi Xu

398 3D Geometry-Aware Deformable Gaussian Splatting for Dynamic 
View Synthesis, Zhicheng Lu, Xiang Guo, Le Hui, Tianrui Chen, 
Min Yang, Xiao Tang, Feng Zhu, Yuchao Dai

399 Boosting Diffusion Models with Moving Average Sampling in 
Frequency Domain, Yuuri Qian, Qi Cai, Yingwei Pan, Yehao Li, 
Ting Yao, Qibin Sun, Tao Mei

400 NoiseCollage: A Layout-Aware Text-to-Image Diffusion Model 
Based on Noise Cropping and Merging, Takahiro Shirakawa, 
Seichi Uchida

401 NeRF On-the-go: Exploiting Uncertainty for Distractor-free NeRFs 
in the Wild, Weining Ren, Zihan Zhu, Boyang Sun, Jiachi Chen, 
Marc Pollefeys, Songyong Peng

402 Using Human Feedback to Fine-tune Diffusion Models without 
Any Reward Model, Kai Yang, Jian Tao, Jiafei Lyu, Chunjiang Ge, 
Jiaxin Chen, WeiHan Shen, Xiaolong Zhu, Xi Li

403 GeneAvatar: Generic Expression-Aware Volumetric Head Avatar 
Editing from a Single Image, Chong Bao, Yinda Zhang, Yuan Li, 
Xuyu Zhang, Bangbang Yang, Hujun Bao, Marc Pollefeys, Guofeng Zhang, Zhaopeng Cui

404 MaskPLAN: Masked Generative Layout Planning from Partial 
Input, Hang Zhang, Anton Savov, Benjamin Dillenburger

405 WOWAIF: Weight Modulation for User Attribution and 
Fingerprinting in Text-to-Image Diffusion Models, Changhoon Kim, Kyle Min, Maitreya Patel, Sheng Chong, 
Yezhou Yang

406 Transcoding Forgery Specificity with Latent Space 
Augmentation for Generalizable Deepfake Detection, Zhiyuan Yan, Yuhao Luo, Siwei Lyu, Qingshan Liu, Baoyuan Wu

407 SCEdit: Efficient and Controllable Image Diffusion Generation via
Skip Connection Editing, Zeyinzi Jiang, Chaojie Mao, Yulin Pan, 
Zhen Han, Jingfeng Zhang

408 CONFORM: Contrast is All You Need for High-Fidelity Text-to-Image 
Diffusion Models, Tun-Han Salih Merai, Enis Simser, 
Federico Tombari, Pinar Yanardag

409 TIZV-Zero: Zero-Shot Image Conditioning for Text-to-Video 
Diffusion Models, Haomiao Ni, Bernhard Egger, Suhas Lohit, 
Anoop Cherian, Ye Wang, Toshiaki Koike-Akino, Sharon X. Huang, 
Tim K. Marks

410 HIVE: Harnessing Human Feedback for Instructional Visual 
Editing, 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

411 Taming Mode Collapse in Score Distillation for Text-to-3D 
Generation, Peihao Wang, Deja Xu, Zhiwen Fan, Dilion Wang, 
Sreyas Mohan, Forrest Iandola, Rakesh Ranjan, Yilei Li, Qiang Liu, 
Zhangyang Wang, Vikas Chandra

412 CoDi: Conditional Diffusion Distillation for Higher-Fidelity and 
Faster Image Generation, Kangyu Mei, Mauricio Delbracio, 
Hossein Talebi, Zhengzhong Tu, Vishal M. Patel, Peyman Milanfar

413 Universal Robustness via Median Randomized Smoothing for Real-
World Super-Resolution, Zakariya Chouawi, Mohamed Tamaazousti

414 ECLIPSE: A Resource-Efficient Text-to-Image Prior for Image 
Generations, Maitreya Patel, Changhoon Kim, Sheng Chong, 
Chitta Baral, Yezhou Yang

415 CAMEL: Causal Motion Enhancement Tailored for Lifting Text-
driven Video Editing, Guwei Zhang, Tianyu Zhang, Guanglin Niu, 
Zichang Tan, Yalong Bai, Qing Yang

416 FreeCustom: Tuning-Free Customized Image Generation for 
Multi-Concept Composition, Ganggui Ding, Canyu Zhao, Wen 
Wang, Zhen Yang, Zide Liu, Hao Chen, Chunhua Shen

417 Amodal Completion via Progressive Mixed Context Diffusion, 
Katherine Xu, Lingzhi Zhang, Jianbo Shi

418 Named Entity Driven Zero-Shot Image Manipulation, Zhida Feng, 
Li Chen, Jing Tian, JiXiang Liu, Shikun Feng

419 Learning Degradation-unaware Representation with Prior-based 
Latent Transformations for Blind Face Restoration, Lixin Xie, 
Csbingbing Zheng, Wen Xue, Le Jiang, Cheng Liu, Si Wu, Hau San Wong

420 AEROBLADE: Training-Free Detection of Latent Diffusion Images 
Using Autocoder Reconstruction Error, Jonas Rickre, Denis Lukovnikov, Asja Fischer

421 VRTouchER: Learning Cross-frame Feature Interdependence 
with Imperfection Flow for Face Retouching in Videos, Wen Xue, 
Le Jiang, Lixin Xie, Si Wu, Yong Xu, Hau San Wong

422 Generative Unlearning for Any Identity, Juwon Seo, Hoon-Hoe Lee, 
Tae-Young Lee, Seungjun Moon, Gyeong-Moon Park

423 Doubly Abductive Counterfactual Inference for Text-based Image 
Editing, Xue Song, Jiequan Cui, Hanwang Zhang, Jingjing Chen, 
Richang Hong, Yu-Gang Jiang

424 Text-conditional Attribute Alignment across Latent Spaces for 3D 
Controllable Face Image Synthesis, Feifan Xu, Rui Li, Si Wu, Yong 
Xu, Hau San Wong

425 Customization Assistant for Text-to-Image Generation, Yufan Zhou, 
Ruiyi Zhang, Xiujiang Gu, Tong Sun

426 Contrastive Denoising Score for Text-guided Latent Diffusion 
Image Editing, Hyelin Nam, Gihyun Kwon, Geon Yeong Park, 
Jong Chul Ye

427 Arbitrary-Style Image Generation and Upsampling using Latent 
Diffusion Model and Implicit Neural Decoder, Jinseok Kim, 
Tae-Kyun Kim

428 VMC: Video Motion Customization using Temporal Attention 
Adaptation for Text-to-Video Diffusion Models, Hyeono Jeong, 
Geon Yeong Park, Jong Chul Ye

429 Visual Layout Composer: Image-Vector Dual Diffusion Model for 
Design Layout Generation, Mohammad Amin Shabani, Zhaowen Wang, Difan Liu, Nanxuan Zhao, Jimei Yang, Yasutaka Furukawa

430 Learning Spatial Adaptation and Temporal Coherence in Diffusion 
Models for Video Super-Resolution, Zhikai Chen, Fuchen Long, 
Zhaofan Qiu, Ting Yao, Wengang Zhou, Jiebo Liu, Tao Mei

431 Open-Vocabulary Attention Maps with Token Optimization for 
Semantic Segmentation in Diffusion Models, Pablo Marcos-Manchón, 
Roberto Alcover-Couso, Juan C. SanMiguel, José M. Martínez
Combining Frame and GOP Embeddings for Neural Video Representation, Jens Eirik Saethre, Roberto Azevedo, Christopher Schroers


Your Student is Better Than Expected: Adaptive Teacher-Student Collaboration for Text-Conditional Diffusion Models, Nikita Starodubcev, Dmitry Baranchuk, Artem Fedorov, Artem Babenko

Mitigating Motion Blur in Neural Radiance Fields with Events and Frames, Marco Cannici, Davide Scaramuzza

Unmixing Before Fusion: A Generalized Paradigm for Multi-Source-based Hyperspectral Image Synthesis, Yang Yu, Erting Pan, Xinya Wang, Yuheng Wu, Xiaoguang Mei, Jiayi Ma


DisCo: Disentangled Control for Realistic Human Dance Generation, Tan Wang, Linjie Li, Kevin Lin, Yuanhao Zhai, Chung-Ching Lin, Zhengyuan Yang, Hanwang Zhang, Zicheng Liu, Lijuan Wang

The Devil is in the Details: StyleFeatureEditor for Detail-Rich StyleGAN Inversion and High Quality Image Editing, Denis Bobkov, Vadim Titov, Aibek Alanov, Dmitry Vetrov

C3: High-Performance and Low-Complexity Neural Compression from a Single Image or Video, Hyunjik Kim, Matthias Bauer, Lucas Theis, Jonathan Richard Schwarz, Emilien Dupont

LightIt: Illumination Modeling and Control for Diffusion Models, Peter Kocsis, Julien Philip, Kalyan Sunkavalli, Matthias Nießner, Yannick Hold-Geoffroy

Rethinking the Spatial Inconsistency in Classifier-Free Diffusion Guidance, Dazhong Shen, Guanglu Song, Zeyue Xue, Fu-Yun Wang, Yu Liu

InitNO: Boosting Text-to-Image Diffusion Models via Initial Noise Optimization, Xiefan Guo, Jinlin Liu, Miaomiao Cui, Jiankai Li, Hongyu Yang, Di Huang

On the Diversity and Realism of Distilled Dataset: An Efficient Dataset Distillation Paradigm, Peng Sun, Bei Shi, Daiwei Yu, Tao Lin

On the Scalability of Diffusion-based Text-to-Image Generation, Hao Li, Yang Zou, Ying Wang, Orchid Majumder, Yusheng Xie, R. Manmatha, Ashwin Swaminathan, Zhuowen Tu, Stefano Ermon, Stefano Soatto

Distilling ODE Solvers of Diffusion Models into Smaller Steps, Sanghwan Kim, Hao Tang, Fisher Yu

ZeroNVS: Zero-Shot 360-Degree View Synthesis from a Single Image, 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

Fixed Point Diffusion Models, Xingjian Bai, Luke Melas-Kyriazi

Gaussian Shell Maps for Efficient 3D Human Generation, Rameen Abdal, Wang Yifan, Zifan Shi, Yinghao Xu, Ryan Po, Zhengfei Kuang, Qifeng Chen, Dit-Yan Yeung, Gordon Wetzstein

Inversion-Free Image Editing with Language-Guided Diffusion Models, Sihan Xu, Yidong Huang, Jiayi Pan, Ziqiao Ma, Joyce Chai

TIGER: Time-Varying Denoising Model for 3D Point Cloud Generation with Diffusion Process, Zhiyuan Ren, Minchul Kim, Feng Liu, Xiaoming Liu

Beyond First-Order Tweedie: Solving Inverse Problems using Latent Diffusion, Litu Rout, Yuja Chen, Abhishek Kumar, Constantine Caramanis, Sanjay Shakkottai, Wen-Sheng Chu

U-VAP: User-specified Visual Appearance Personalization via Decoupled Self Augmentation, You Wu, Kuan Liu, Xiaoyue Mi, Fan Tang, Juan Cao, Jintao Li

19:00 - 21:00 SOCIAL - Diversity and Inclusion for Everyone (Summit Garden Terrace) **Reservation Required
Thursday, June 20

**Registration / Badge Pickup** (Summit Lobby)
9:00 - 12:00

**Press Room** (Summit 340)
9:00 - 10:30

**Mother's Room** (Summit 341-adjacent and Summit 441-adjacent)
9:00 - 10:30

**Prayer or Quiet Room** (Summit 326)
9:30 - 9:50

**Breakfast** (Summit ExHall 1-2)
8:30 - 10:30

**SOCIAL - How to Know Your True Market Value as an AI Researcher** (Summit Terrace Suite)
**Reservation Required**

**Orals 3A: 3D from Single View** (Summit Ballroom)
9:00 - 10:30
1. Repurposing Diffusion-Based Image Generators for Monocular Depth Estimation, Bingxin Ke, Anton Obukhov, Shengyu Huang, Nando Metzger, Rodrigo Case Daudt, Konrad Schindler
2. EscherNet: A Generative Model for Scalable View Synthesis, Xin Kong, Shikun Liu, Xiaoyang Lyu, Marwan Taher, Xiaojuan Qi, Andrew J. Davison
3. WAL3TD: Generating Realistic Training Data from Time-Lapse Imagery for Reconstructing Dynamic Objects Under Occlusion, Khiem Vuong, N Dinesh Reddy, Robert Tamburo, Swinvasa G. Narasimhan
5. Rethinking Inductive Biases for Surface Normal Estimation, Gwangbin Bae, Andrew J. Davison

**Orals 3B: Vision, Language, and Reasoning** (Summit Flex Hall AB)
9:00 - 10:30
1. Comparing the Decision-Making Mechanisms by Transformers and CNNs via Explanation Methods, Mingqiang Jiang, Saeed Khorraram, Li Fuxin
2. MMMU: A Massive Multi-discipline Multimodal Understanding and Reasoning Benchmark for Expert AGI, Xiang Yue, Yuansheng Ni, Kai Zhang, Tianyu Zheng, Ruoci Liu, Ge Zhang, Samuel Stevens, Dongfu Jiang, Weiming Ren, Yuxuan Sun, Cong Wei, Botao Yu, Ruibin Yuan, Renliang Sun, Ming Yin, Boyuan Zheng, Zhenhu Zhang, Yibo Liu, Wenhao Huang, Huan Sun, Yu Su, Wenhu Chen
3. Eyes Wide Shut? Exploring the Visual Shortcomings of Multimodal LLMs, Shengbang Tong, Zhuang Liu, Yuxiang Zhai, Yi Ma, Yann LeCun, Saining Xie
4. LISA: Reasoning Segmentation via Large Language Model, Xin Lai, Zhuotao Tian, Yukang Chen, Yanwei Li, Yuhiu Yuan, Shu Liu, Jiayu Jia

**Orals 3C: Medical and Physics-Based Vision** (Summit Flex Hall C)
9:00 - 10:30
1. EventPS: Real-Time Photometric Stereo Using an Event Camera, Bohan Yu, Jieqi Ren, Jin Han, Feishi Wang, Jinxiu Liang, Boxin Shi
2. EvD3G: Event-guided Direct and Global Components Separation, Xinyu Zhou, Peiqi Duan, Boyu Li, Chu Zhou, Chao Xu, Boxin Shi
3. MemSAM: Taming Segment Anything Model for Echocardiography Video Segmentation, Xiaolong Deng, Huisi Wu, Runhao Zeng, Jing Qin
5. Correlation-aware Coarse-to-fine MLPs for Deformable Medical Image Registration, Mingyuan Meng, Dagan Feng, Lei Ji, Jinman Kim

**Poster Setup** (Arch 4E)
10:00 - 10:30

**Expo Track**
10:30 - 11:30
Andrea Gagliano
Senior Director, AI/ML at Getty Images (Arch 4F)

**Orals 3B: Vision, Language, and Reasoning**
10:30 - 12:00
**Poster Session 3 & Exhibit Hall** (Arch 4A-E)
2. CityDreamer: Compositional Generative Model of Unbounded 3D Cities, Haozhe Xie, Zhaochi Chen, Fangzhou Hong, Ziwei Liu
3. 6D-Diff: A Keypoint Diffusion Framework for 6D Object Pose Estimation, Li Xu, Haoxuan Qu, Yujun Cai, Jun Liu
5. Repurposing Diffusion-Based Image Generators for Monocular Depth Estimation, Bingxin Ke, Anton Obukhov, Shengyu Huang, Nando Metzger, Rodrigo Case Daudt, Konrad Schindler
7. WorDepth: Variational Language Prior for Monocular Depth Estimation, Ziyao Zeng, Daniel Wang, Fengyu Yang, Hyoungseob Park, Stefano Soatto, Dong Lao, Alex Wong
8. Free3D: Consistent Novel View Synthesis without 3D Representation, Chunxia Zheng, Andrew Vedaldi
10. 3DFIRES: Few Image 3D Reconstruction for Scenes with Hidden Surfaces, Linyi Jin, Nileshe Kularkar, David F. Fouhey
11. Learning the 3D Fauna of the Web, Zhuizhe Li, Dor Litvak, Ruining Li, Yunzhi Zhang, Tomasz Jakab, Christian Rupprecht, Shangzhe Wu, Andrea Vedaldi, Jiajun Wu
12. Bilateral Propagation Network for Depth Completion, Jie Tang, Fei-Feng Tian, Boshu An, Jian Li, Ping Tan
14. EpiDiff: Enhancing Multi-View Synthesis via Localized Epipolar-Constrained Diffusion, Zehuan Huang, Hao Wen, Junting Dong, Yaohui Wang, Yangguang Li, Xinyuan Chen, Yan-Pei Cao, Ding Liang, Yu Qiao, Bo Dai, Lu Sheng
15. Doodle Your 3D: From Abstract Freehand Sketches to Precise 3D Shapes, Hrhmishrat Bandyopadhyay, Subhadip Koley, Ayan Das, Ayan Kumar Bhunia, Aneesha Sain, Pinaki Nath Chowdhury, Tao Xiang, Yi-Zhe Song
17. CNC-Net: Self-Supervised Learning for CNC Machining Operations, Mohnsen Yavartanoo, Sangmin Hong, Reyhaneh Neshatavar, Kyoung Mu Lee
18. Reconstructing Hands in 3D with Transformers, Georgios Pavlakos, Dandan Shan, Ilija Radosavovic, Angjoo Kanazawa, David Fouhey, Jitendra Malik
19. Boosting Self-Supervision for Single-View Scene Completion via Knowledge Distillation, Keonhee Han, Dominik Muhe, Felix Wimbauer, Daniel Cremers
22. ViewFusion: Towards Multi-View Consistency via Interpolated Denoising, Xianghui Yang, Yan Zuo, Sameera Ramasinghe, Loris Bazzani, Gil Avraham, Anton van den Hengel
24. Consistent3D: Towards Consistent High-Fidelity Text-to-3D Generation with Deterministic Sampling Prior, Zike Wu, Pan Zhou, Xuanyu Yi, Xiaodong Yuan, Hanwang Zhang
GigaPose: Fast and Robust Novel Object Pose Estimation via One Correspondence, Van Nguyen Nguyen, Thibault Groueix, Mathieu Salzmann, Vincent Lebart

RichDreamer: A Generalizable Normal-Depth Diffusion Model for Detail Richness in Text-to-3D, Lingteng Qiu, Guanying Chen, Xiaodong Gu, Qi Zuo, Mutian Xu, Yushuang Wu, Weihao Yuan, Zilong Dong, Liefeng Bo, Xiaoguang Han

Elite360D: Towards Efficient 360 Depth Estimation via Semantic-and Distance-Aware Bi-Projection Fusion, Hao Ai, Lin Wang

SIFU: Side-view Conditioned Implicit Function for Real-world Usable Clothed Human Reconstruction, Zechuan Zhang, Zhibin Lin, Zexuan Zhang, Ziyang Li, Yiyang Yang, Yanghao Wu, Zhaoyang Liu

Diffusion Time-step Curriculum for One Image to 3D Generation, Xuan Yu, Zike Wu, Qingshan Xu, Pan Zhou, Joo-Hwee Lim, Hanwang Zhang

SecondPose: SE(3)-Consistent Dual-Stream Feature Fusion for Category-Level Pose Estimation, Yamei Chen, Yan Di, Guangyao Zhai, Fabian Manhardt, Chenyangguang Zhang, Ruida Zhang, Federico Tombari, Nassir Navab, Benjamin Busam

Wonder3D: Single Image to 3D using Cross-Domain Diffusion, Xiaoxiao Long, Yuan-Chen Guo, Cheng Lin, Yuan Liu, Zhiyang Dou, Lingjie Liu, Yuexin Ma, Song-Hai Zhang, Marc Habermann, Christian Theobalt, Wenping Wang

En3D: An Enhanced Generative Model for Sculpting 3D Humans from 2D Synthetic Data, Yifang Men, Biwen Lei, Yuan Yao, Miaomiao Cui, Zhouhui Lian, Xuanxiong Xie, Minyi Li, Chao Peng, Guangning Li, Ruida Zhang, Fabian Manhardt, Bowen Fu, Federico Tombari, Xiangyang Ji

MOHO: Learning Single-view Hand-held Object Reconstruction with Multi-view Occlusion-Aware Supervision, Chenyangguang Zhang, Guanlong Jiao, Yan Di, Gu Wang, Ziqing Huang, Ruida Zhang, Fabian Manhardt, Bowen Fu, Federico Tombari, Xiangyang Ji


PatchFusion: An End-to-End Tile-Based Framework for High-Resolution Monocular Depth Estimation, Zhenyu Li, Shariq Farooq Bhat, Peter Wonka

SPAD: Spatially Aware Multi-View Diffusers, Yash Kant, Aliaksandr Sirahin, Ziyi Wu, Michael Vasilevskiy, Guoqiang Qian, Jian Ren, Riza Alp Guler, Bernard Ghanem, Sergey Tulyakov, Igor Gilitschenski


PointInk: Resolution-Invariant Point Diffusion Models, Zixuan Huang, Justin Johnson, Shoubhik Debnath, James M. Rehg, Chao-Yuan Wu

ZeroShape: Regression-based Zero-shot Shape Reconstruction, Zixuan Huang, Stefan Stojanov, Anh Thai, Varun Jampani, James M. Rehg

One-2-3-45+: Fast Single Image to 3D Objects with Consistent Multi-View Generation and 3D Diffusion, Minghua Liu, Ruoxi Shi, Linghao Chen, Zhourang Zhang, Chao Xu, Xinyue Wei, Hansheng Chen, Chong Zeng, Jiayuan Gu, Hao Su

ConTekHuman: Free-View Rendering of Human from a Single Image with Texture-Consistent Synthesis, Xiqian Gao, Xiaoyu Li, Chaopeng Zhang, Zhi Qiang, Yanpei Cao, Mingkui Tan, Long Quan MatchU: Matching Unseen Objects for 6D Pose Estimation from RGB-D Images, Junwen Huang, Hao Yu, Kuan-Ting Yu, Nassir Navab, Slobodan Ilic, Benjamin Busam

UniDepth: Universal Monocular Depth Estimation, Luigi Pinzelli, Yung-Hsu Yang, Christos Sarakidis, Mattia Segu, Siyuan Li, Luc Van Gool, Fisher Yu

G-NeRF: Geometry-enhanced Novel View Synthesis from Single-View Images, Zixiong Huang, Qi Chen, Libo Sun, Yifan Yang, Naizhou Wang, Qi Wu, Mingkui Tan

3DToonify: Creating Your High-Fidelity 3D Stylized Avatar Easily from 2D Portrait Images, Yifang Men, Hanxi Liu, Yuan Yao, Miaomiao Cui, Xuanxiong Xie, Zhouhui Lian

Adaptive Fusion of Single-view and Multi-view Depth for Autonomous Driving, Junda Cheng, Wei Yi, Kaixuan Wang, Xiaochi Chen, Shijie Wang, Xin Yang

HiPose: Hierarchical Binary Surface Encoding and Correspondence Pruning for RGB-D 6DoF Object Pose Estimation, Yongliang Lin, Yongzhi Su, Praveen Nathan, Sandeep Inugant, Yan Di, Martin Sundermeyer, Fabian Manhardt, Didier Stricker, Jason Rambach, Yu Zhang

HandBooster: Boosting 3D Hand-Mesh Reconstruction by Conditional Sampling of Hand-Object Interactions, Hao Xu, Haipeng Li, Yinqiao Wang, Shuaicheng Liu, Chi-Wing Fu

3D-SceneDreamer: Text-Driven 3D-Consistent Scene Generation, Songchun Zhang, Yibo Zhang, Quan Zheng, Rui Ma, Wei Hua, Hujun Bao, Weixi Xu, Changqing Zu

NVIST: In the Wild New View Synthesis from a Single Image with Transformers, Wonbong Jang, Lourdes Agapito

CAD: Photorealistic 3D Generation via Adversarial Distillation, Ziyu Wan, Despoina Paschalidou, Ian Huang, Hongyu Liu, Bokui Shen, Xiaoyu Xiang, Jing Liao, Leonidas Guibas


Joint Reconstruction of 3D Human and Object via Contact-Based Refinement Transformer, Hyungjin Nam, Daniel Sungho Jung, Gyeongsik Moon, Kyoung Mu Lee

Sculpt3D: Multi-View Consistent Text-to-3D Generation with Sparse 3D Prior, Cheng Chen, Xiaofeng Yang, Fan Yang, Chengzeng Feng, Zhoujie Fu, Chuan-Sheng Foo, Guosheng Lin, Fayao Liu

Object Pose Estimation via the Aggregation of Diffusion Features, Tianfu Wang, Guosheng Hu, Hongguang Wang

MonoCD: Monocular 3D Object Detection with Complementary Depths, Longfei Yan, Pei Yan, Shengzhou Xiong, Xuyang Xiang, Yihua Tan

MultiDiff: Consistent Novel View Synthesis from a Single Image, Norman Müller, Katja Schwarz, Barbara Rössle, Lorenzo Porzi, Samuel Rota Bulò, Matthias Nießner, Peter Kontschieder

SeaBird: Segmentation in Bird’s View with Dice Loss Improves Monocular 3D Detection of Large Objects, Abhinav Kumar, Yuliang Lin, Xiaoyu Li, Xin Yang, Ruida Zhang, Xiaogang Zhuang, Xiaoyu Li, Tao Li, Xiaoming Liu

Learning Occupancy for Monocular 3D Object Detection, Liang Peng, Junkai Xu, Haoaren Cheng, Zicheng Yang, Xiaopei Wu, Wei Qian, Wenxiao Wang, Boxi Wu, Deng Cai

NeRFDeformer: NeRF Transformation from a Single View via 3D Scene Flows, Zhenggang Tang, Zhongzheng Ren, Xiaoming Zhao, Bowen Wen, Jonathan Tremblay, Stan Birchfield, Alexander Schwing

R-Cyclic Diffuser: Reductive and Cyclic Latent Diffusion for 3D Clothed Human Digitalization, Kennard Yanting Chan, Fayao Liu, Guosheng Lin, Chuan Sheng Foo, Weisi Lin

Unleashing Network Potentials for Semantic Scene Completion, Fengyun Wang, Qianru Sun, Dong Zhang, Jinhui Tang

Triplane Meets Gaussian Splatting: Fast and Generalizable Single-View 3D Reconstruction with Transformers, Zi-Xin Zou, Zhipeng Yu, Yuan-Chen Guo, Yangguang Li, Ding Liang, Yan-pei Cao, Song-Hai Zhang

VOODOO 3D: Volumetric Portrait Disentanglement For One-Shot 3D Head Reenactment, Phong Tran, Egor Zakharov, Long-Nhat Ho, Anh Tuan Tran, Liwen Hu, Hao Li

Compressed 3D Splatting for Accelerated Novel View Synthesis, Simon Niedermayer, Josef Stumpfegger, Rüdiger Westermann

Morphable Diffusion: 3D-Consistent Diffusion for Single-image Avatar Creation, Xiyi Chen, Marko Mihajlovic, Shaofei Wang, Sergey Prokudin, Siyu Yang

Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data, Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, Hengshuang Zhao

SAOR: Single-View Articulated Object Reconstruction, Mehmet Aygun, Olisn Mac Aodha

EscherNet: A Generative Model for Scalable View Synthesis, Xin Kong, Shikun Liu, Xiaoyang Lyu, Marwan Taher, Xiaojian Qi, Andrew J. Davison

HOISDF: Constraining 3D Hand-Object Pose Estimation with Global Signed Distance Fields, Haozhe Qi, Chen Zhao, Mathieu Salzmann, Alexander Mathis

Diffusion-driven GAN Inversion for Multi-Modal Face Image
Generation, Ji hyun Kim, Changjae Oh, Hoseok Do, So hyun Kim, Kwang hoon Sohn
72 Novel View Synthesis with View-Dependent Effects from a Single Image, Juan Luis Gonzalez Bello, Munchurl Kim
73 Weakly-Supervised Emotion Transition Learning for Diverse 3D Co-speech Gesture Generation, Xingqun Qi, Jiahao Pan, Peng Li, Ruibin Yuan, Xiaowei Chi, Mengfei Li, Wenhan Luo, Wei Xue, Shanghang Zhang, Qifeng Liu, Yike Guo
74 Seq2Reg: Differentiable 2D Segmentation to 3D Regression Rendering for 360 Room Layout Reconstruction, Cheng Sun, Wei-En Tai, Yu-Lin Shih, Kuan-Wei Chen, Yong-Jing Syu, Kent Selwyn The, Yu-Chiang Frank Wang, Hwann-Tzong Chen
75 Mining Supervision for Dynamic Regions in Self-Supervised Monocular Depth Estimation, Hoang Chuong Nguyen, Tianyu Wang, Jose M. Alvarez, Miaomiao Liao
76 WALT3D: Generating Realistic Training Data from Time-Lapse Imagery for reconstructing Dynamic Objects Under Occlusion, Khiem Vuong, N Dinesh Reddy, Robert Tamburo, Srinivasa G. Narasimhan
77 DiffPortrait3D: Controllable Diffusion for Zero-Shot Portrait View Synthesis, Yuming Gu, Hongyi Xu, You Xie, Guoxian Song, Yichun Shi, Di Chang, Jing Yang, Linjie Luo
78 3D-LFM: Lifting Foundation Model, Mosam Dabhi, Lásló A. Jeni, Simon Lucey
79 MRC-Net: 6-DoF Pose Estimation with MultiScale Residual Correlation, Yuelong Li, Yafei Mao, Raja Bala, Sunil Hadap
80 DiffusionGAN3D: Boosting Text-guided 3D Generation and Domain Adaptation by Combining 3D GANs and Diffusion Priors, Biwen Lei, Kai Yu, Mengyang Feng, Miaomiao Cui, Xuan song Xie
81 VS: Reconstructing Clothed 3D Human from Single Image via Vertex Shift, Leyuan Liu, Yuhang Li, Yunqi Gao, Changxin Gao, Youyuan Liu, Jingyin Chen
82 Weakly Supervised Monocular 3D Detection with a Single-View Image, Xueying Jang, Sheng Jin, Lewei Lu, Xiaojing Yang, Shijian Lu
83 From-Ground-To-Objects: Coarse-to-Fine Self-supervised Monocular Depth Estimation of Dynamic Objects with Ground Contact Prior, Jae Ho Moon, Juan Luis Gonzalez Bello, Byeongjun Kwon, Munchurl Kim
84 Gated Fields: Learning Scene Reconstruction from Gated Videos, Andrea Ramazzina, Stefanie Walz, Pragyan Dahal, Mario Bijelic, Felix Heide
85 SCINeRF: Neural Radiance Fields from a Snapshot Compressive Image, Yunhao Li, Xiaodong Wang, Ping Wang, Xin Yuan, Peidong Liu
86 Diffusion-FOF: Single-View Clothed Human Reconstruction via Diffusion-Based Fourier Occupancy Field, Yuanzhen Li, Fei Luo, Chuxiao Xia
87 Instance-aware Contrastive Learning for Occluded Human Mesh Reconstruction, Mi-Gyeong Gwon, Gi-Mun Um, Won-Sik Cheong, Wonjun Kim
88 IB-SLAM: Learning Image-Based Depth Fusion for Generalizable SLAM, Minghao Yin, Shangzhe Wu, Kai Han
89 HarmonyView: Harmonizing Consistency and Diversity in One-Image-to-3D, Sangmin Woo, Byeongjun Park, Hyoujun Go, Jin-Young Kim, Changick Kim
90 UV-IDM: Identity-Conditioned Latent Diffusion Model for Face UV-Textrue Generation, Hong Li, Yutang Feng, Song Xue, Xuhui Liu, Bohan Zeng, Shanlin Li, Boyu Liu, Jianzhuang Liu, Shumin Han, Baochang Zhang
91 Attrihuman-3D: Editable 3D Human Avatar Generation with Attribute Decomposition and Indexing, Fan Yang, Tianyi Chen, Xiaosheng He, Zhongang Cai, Lei Yang, Si Wu, Guosheng Lin
92 Mind The Edge: Refining Depth Edges in Sparsely-Supervised Monocular Depth Estimation, Lior Talker, Aviad Cohen, Erez Yosef, Alexandar Dana, Michael Dinerstein
93 3Diffection: 3D Object Detection with Geometry-Aware Diffusion Features, Chenfeng Xu, Huan Ling, Sanja Fidler, Or Litany
94 Bayesian Diffusion Models for 3D Shape Reconstruction, Haiyang Xu, Yu Lei, Zeyuan Chen, Xiang Zhang, Yue Zhao, Yilin Wang, Zhiwen Tu
95 Rethinking Inductive Biases for Surface Normal Estimation, Kwangbin Bae, Andrew J. Davison
96 LaneCPP: Continuous 3D Lane Detection using Physical Prriors, Maximilian Pittner, Joel Janai, Alexandru M. Condurache
97 Enhancing 3D Fidelity of Text-to-3D using Cross-View Correspondences, Seungwook Kim, Keiji Li, Xueqing Deng, Yichun Shi, Min Su Cho, Peng Wang
98 MonoDiff: Monocular 3D Object Detection and Pose Estimation with Diffusion Models, Yisuru Ranasinghe, Deepti Hegde, Vishal M. Patel
99 HiLo: Detailed and Robust 3D Clothed Human Reconstruction with High-and Low-Frequency Information of Parametric Models, Yifan Yang, Dong Liu, Shuhai Zhang, Zeshuai Deng, Zixiong Huang, Mingkui Tan
100 MPOD23: One Image to 3D Content Generation Using Mask-enhanced Progressive Outline-to-Detail Optimization, Jimin Xu, Tianbao Wang, Tao Jin, Shengyu Zhang, Dongjie Fu, Ze Wang, Jianjiang Lyu, Chengfei Lü, Chaoque Niu, Zhou Yu, Zhou Zhao, Fei Wu
101 GeoReF: Geometric Alignment Across Shape Variation for Category-level Object Pose Refinement, Linfang Zheng, Tze Ho Elden Tse, Chen Wang, Yinghan Sun, Hua Chen, Ales Leonardis, Wei Zhang, Huying Jin Chang
102 Unsupervised 3D Structure Inference from Category-Specific Image Collections, Weikang Wang, Dongliang Cao, Florian Bernard
103 Neural Parametric Gaussians for Monocular Non-Rigid Object Reconstruction, Devikalyan Das, Christopher Wewer, Raza Yunus, Eddy Ilg, Jan Eric Lenssen
104 BiIT: Bi-directional Texture Reconstruction of Interacting Two Hands from a Single Image, Minjie Kim, Tae-Kyun Kim
105 DeCoTR: Enhancing Depth Completion with 2D and 3D Atentions, Yuxun Xiao, Manish Kumar Singh, Hong Cai, Faith Porikli
106 MonoNPHM: Dynamic Head Reconstruction from Monocular Videos, Simon Giebenhain, Tobias Kirschstein, Markus Georgopoulos, Martin Rünz, Lourdes Agapito, Matthias Nießner
107 Fakelnversion: Learning to Detect Images from Unseen Text-to-Image Models by Inverting Stable Diffusion, George Cazenevete, Avesheh Sud, Thomas Leung, Ben Usman
108 Forgery-aware Adaptive Transformer for Generalizable Synthetic Image Detection, Huan Liu, Zichang Tian, Chuangchuan Tan, Yunchao Wei, Jingdong Wang, Yao Zhao
109 Towards Modern Image Manipulation Localization: A Large-Scale Dataset and Novel Methods, Chenfan Qu, Yiwu Zhong, Chongyu Liu, Gui ta Xu, Dezhi Peng, Fengjun Guo, Lianwen Jin
110 ManFPT: Defining and Analyzing Fingerprints of Generative Models, Hae Jin Song, Mahyar Khayatkhoee, Wael AbdAlmageed
111 ProMark: Proactive Diffusion Watermarking for Causal Attribution, Vishal Ansani, John Collomosse, Tu Bui, Xiaoming Liu, Shrutti Agarwal
112 CGI-DM: Digital Copyright Authentication for Diffusion Models via Contrasting Gradient Inversion, Xiaoyu Wu, Yang Hua, Chumeng Liang, Jiaru Zhang, Hao Wang, Tao Song, Haibing Guan
113 SCOFT: Self-Contrastive Fine-Tuning for Equitable Image Generation, Zhixuan Liu, Peter Schaldenbrand, Beverley-Claire Okogwu, Wenxuan Peng, Youngsik Yun, Andrew Hundt, Jihee Kim, Jean Oh
114 Would Deep Generative Models Amplify Bias in Future Models?, Tianwei Chen, Yusuke Hirota, Mayu Otani, Noa Garcia, Yuta Nakashima
115 Training Diffusion Models Towards Diverse Image Generation with Reinforcement Learning, Zichen Miao, Jiang Wang, Ze Wang, Zhengyuan Liang, Liuan Wang, Qiang Qiu, Zicheng Liu
116 Consistency and Uncertainty: Identifying Unreliable Responses from Black-Box Vision-Language Models for Selective Visual Question Answering, Zaid Khan, Yun Fu
118
ToonerGAN: Reinforcing GANs for Obfuscating Automated Facial Indexing, Kartik Thakral, Shashikant Prasad, Sutti Aswani, Mayank Vatsala, Richa Singh, Yannic Neuhaus, Matthias Hein

119 MCPNet: An Interpretable Classifier via Multi-Level Concept Prototypes, Bor-Shun Wang, Chien-Yi Wang, Wei-Chen Chu

120 Visual Concept Connectome (VCC): Open World Concept

121 Discovery and their Interlayer Connections in Deep Models, Matthew Kowal, Richard P. Wildes, Konstantinos G. Derpanis


123 Token Transformation Matters: Towards Faithful Post-hoc Explanation for Vision Transformer, Junyi Wu, Bin Duan, Weitai Kang, Hao Tang, Yan Yan

124 Comparing the Decision-Making Mechanisms by Transformers and CNNs via Explanation Methods, Mingqi Jiang, Saeed Khorsam, Li Puxin

125 On the Faithfulness of Vision Transformer Explanations, Junyi Wu, Weitai Kang, Hao Tang, Yuan Hong, Yan Yan

126 Understanding Video Transformers via Universal Concept

127 Discovery, Matthew Kowal, Achal Dave, Rares Ambrus, Adrien Gaidon, Konstantinos G. Derpanis, Pavel Tokmakov

128 Explaining the Implicit Neural Canvas: Connecting Pixels to Neurons by Tracing their Contributions, Namitha Padmanabhan, Matthew Gwilliam, Pulkit Kumar, Shishira R Maiya, Max Ehrlich, Abhinav Shrivastava

129 WWW: A Unified Framework for Explaining What Where and Why of Neural Networks by Interpretation of Neuron Concepts, Yong Hyun Ahn, Hyeon Bae Kim, Seong Tae Kim

130 HDQM: Holographic Feature Decomposition Using Quantum Algorithms, Prathvish Prasanth Povul, Zhouwen Mou, Mohnen Imani

131 SLICE: Stabilized LIME for Consistent Explanations for Image Classification, Revoti Prasad Bora, Philipp Thörrst, Raymond Veldhuis, Raghavendra Ramachandra, Kiran Raju

132 What Sketch Explainability Really Means for Downstream Tasks?, HMrishav Bandyopadhyay, Pinaki Nath Chowdhury, Ayan Kumar Bhunia, Aneeshan Sain, Tao Xiang, Yi-Zhe Song

133 Structured Gradient-based Interpretations via Norm-Regularized Adversarial Training, Shizhan Gong, Qi Dou, Farzan Farnia

134 Learning Triangular Distribution in Visual World, Ping Chen, Xingpeng Zhang, Chengtao Zhou, Dichao Fan, Peng Tu, Le Zhang, Yanlin Qian

135 Incremental Residual Concept Bottleneck Models, Chenning Zhang, Shijl Zhou, Hengyuan Zhang, Xirzhe Ni, Yujie Yang, Yuwang Wang

136 Uncertainty Visualization via Low-Dimensional Posterior Projections, Omor Yair, Elias Nehme, Toner Michaeli

137 Epistemic Uncertainty Quantification For Pre-Trained Neural Networks, Hanjing Wang, Qiang Ji

138 Interpretable Measures of Conceptual Similarity by Complexity-Constrained Descriptive Auto-Encoding, Alessandro Achille, Greg Ver Steeg, Tian Yu Liu, Matthew Trager, Carson Klingenberg, Stefano Soatto

139 CAPE: CAM as a Probabilistic Ensemble for Enhanced DNN Interpretation, Townim Faisal Chowdhury, Kewon Liao, Vu Minh Hieu Phan, Minh-Son To, Yutong Xie, Kevin Hung, David Ross, Anton van den Hengel, Johan W. Verjans, Zhibin Liao

140 Discovering and Mitigating Visual Biases through Keyword Explanation, Younghyun Kim, Sangwoo Mo, Minkyu Kim, Kyungmin Lee, Jaehe Lee, Jinwoo Shin

141 DIG-IN: Diffusion Guidance for Investigating Networks - Uncovering Classifier Differences Neuron Visualisations and Visual Counterfactual Explanations, Maximilian Augustin, Yannic Neuhaus, Matthias Hein

142 Cross-Dimension Affinity Distillation for 3D EM Neuron Segmentation, Xiaoyu Liu, Miaoamiao Cai, Yindra Chen, Yueyi Zhang, Te Shi, Ruobing Zhang, Xuejin Chen, Zhiewon Shin

143 Continual Self-supervised Learning: Towards Universal Multi-modal Medical Data Representation Learning, Yiwen Ye, Yutong Xie, Jianpeng Zhang, Ziyang Chen, Qi Wu, Yong Xia

144 A Unified Framework for Microscopy Defocus Deblurring with Multi-Branch Transformer and Contrastive Learning, Yuelin Zhang, Pengyu Zheng, Wan Yan, Chengyu Fang, Shinh Shing Cheng

145 CARZero: Cross-Attention Alignment for Radiology Zero-Step Classification, Haoran Lai, Qingsong Yao, Zhihang Jiang, Rongsheng Wang, Zhiyang He, Xiaodong Tao, S. Kevin Zhou

146 Towards Generalizable Tumor Synthesis, Qi Chen, Xiaoxi Chen, Haorui Song, Zhiwei Xiong, Alan Yuille, Chen Wei, Zongwei Zhou

147 Tyche: Stochastic In-Context Learning for Medical Image Segmentation

148 Segmentation, Marianne Rakic, Hallea E. Wong, Jose Javier Gonzalez Ortiz, Beth A. Cinimi, John V. Guttig, Adrian V. Dalca


150 Each Test Image Deserves A Specific Prompt: Continual Test-Time Adaptation for 2D Medical Image Segmentation, Ziyang Yang, Yongsheng Pan, Yiwen Ye, Mengkang Lu, Yong Xia

151 Training Like a Medical Resident: Context-Prior Learning Toward Universal Medical Image Segmentation, Yunhe Gao

152 C2RV: Cross-Regional and Cross-View Learning for Sparse-View CBCT Reconstruction, Yiqun Lin, Jiewen Yang, Huihang Li, Xinxing Ding, Wei Zhao, Xiaoming Li

153 Modality-Agnostic Structural Image Representation Learning for Deformable Multi-Modality Medical Image Registration, Tony C. W. Mok, Zi Li, Yunhao Bai, Jianpeng Zhang, Wei Liu, Yan-Jie Zhou, Ye Ke, Dakai Jin, Yu Shi, Xiaoli Yin, Le Lu, Ling Zhang

154 SI-MIL: Taming Deep MIL for Self-Interpretability in Gigapixel Histopathology, Sarthak Kapse, Pushpak Pari, Srijan Das, Jingwei Zhang, Chao Chen, Maria Kavakopoulos, Joel Saltz, Dimitris Saramaris, Rajarsi R. Gupta, Prateek Prasanna

155 Bootstrapping Chest CT Image Understanding by Distilling Knowledge from X-ray Expert Models, Weiwai Cao, Jianpeng Zhang, Yingda Xia, Tony C. W. Mok, Zi Li, Xianghua Ye, Le Lu, Jian Zheng, Xuyang Tang, Ling Zhang

156 ViLa-MIL: Dual-scale Vision-Language Multiple Instance Learning for Whole Slide Image Classification, Jiangbo Shi, Chen Li, Tieliang Gong, Yefeng Zheng, Huazhu Fu

157 Virtual Immunohistochemistry Staining for Histological Images Assisted by Weakly-supervised Learning, Jiayang Dong, Shenzhen Huang, Xi Li, Junjun Jiang, Xiaopeng Fan, Yongbing Zhang

158 Representing Part-Whole Hierarchies in Foundation Models by Learning Localizability Composability and Decomposability from Anatomy via Self Supervision, Mohammad Reza Hoseinizadeh Tahei, Michael B. Gotway, Jianming Liang

159 XFibrosis: Explicit Vessel-Fiber Modeling for Fibrosis Staging from Liver Pathology Images, Chong Yin, Siqi Liu, Fei Lyu, Jiahao Lu, Sune Darkner, Vincent Wai-Sun Wong, Pong C. Yuen

160 Prompting Vision Foundation Models for Pathology Image Analysis, Chong Yin, Siqi Liu, Kajiyah Zhou, Vincent Wai-Sun Wong, Pong C. Yuen

161 One-Prompt to Segment All Medical Images, Junke Wu, Min Xu

162 Learning Large-Factor EM Image Super-Resolution with Generative Priors, Jiateng Shou, Zeyu Xiao, Shiyu Deng, Wei Huang, Peiyao Shi, Ruobing Zhang, Zhiwei Xiong, Feng Wu

163 Dynamic Graph Representation with Knowledge-aware Attention for Histopathology Whole Slide Image Analysis, Jiawen Li, Yuxuan Chen, Hongbo Chu, Qihe Sun, Tian Guan, Anjia Han, Yonghong He

164 MindBridge: A Cross-Subject Brain Decoding Framework, Shizun Wang, Songhua Liu, Zhenxin Tan, Xinchao Wang

165 Towards Generalizable Tumor Synthesis, Shan Yu, Zhiwei Xiong, Chao Chen, Xiaopeng Fan, Yongbing Zhang

166 Adapting Video-Language Models for Generalizable Anomaly Detection in Medical Images, Chaoqin Huang, Aofan Jiang, Weizhong Zhu, Weisheng Xu, Hongwei Li, Yanjun Gao, Qingchun Zhang, Shouyi Guo, Aihua Wang, Shixing Li, Xiaoyu Shen, Zihao Li, Zhiwei Xiong, Yonghong He, Zhe Sun, Jie Li, Wei Peng, Feng Wu, Zhiwei Xiong, Yongbing Zhang, Chao Chen, Ji Yang, Xiaopeng Fan, Yongbing Zhang
Jinghao Feng, Ya Zhang, Xinchao Wang, Yanfeng Wang

ZePT: Zero-Shot Pan-Tumor Segmentation via Query-Disentangling and Self-Prompting, Yankai Jiang, Zhongzhen Huang, Rongzhao Zhang, Xianfeng Wang, Shouting Zhang

MemSAM: Taming Segment Anything Model for Echocardiography

Video Segmentation, Xiaolong Deng, Huisi Wu, Runhao Zeng, Jing Qin

Generalizable Whole Slide Image Classification with Fine-Grained Visual-Semantic Interaction, Hao Li, Ying Chen, Yifei Chen, Rongshan Yu, Wenxian Yang, Liangsheng Wang, Bowen Ding, Yuchen Han

Incremental Nuclei Segmentation from Histopathological Images via Future-class Awareness and Compatibility-inspired Distillation, Huyong Wang, Huisi Wu, Jing Qin

PH-Net: Semi-Supervised Breast Lesion Segmentation via Patch-wise Hardness, Siyao Jiang, Huisi Wu, Junyang Chen, Qin Zhang, Jing Qin

ToNNO: Tomographic Reconstruction of a Neural Network's Output for Weakly Supervised Segmentation of 3D Medical Images, Marius Schmidt-Mengin, Alexis Benichou, Shibeshi Belachew, Nikos Komodakis, Nikos Paragios

Think Twice Before Selection: Federated Evidential Active Learning for Medical Image Analysis with Domain Shifts, Jiayi Chen, Benteng Ma, Hengfei Cui, Yong Xia


MicroDiffusion: Implicit Representation-Guided Diffusion for Multimodal Reconstruction from Limited 2D Microscopy Projections, Mude Hui, Zhaoh Wei, Haozhi Yu, Fei Xia, Yuxin Zhou

Diversified and Personalized Multi-rater Medical Image Segmentation, Yicheng Wu, Xiangde Luo, Zhe Xu, Xiaoxing Guo, Lie Ju, Zongyuan Ge, Wenjun Liao, Jianfei Cai

Modality-agnostic Domain Generalizable Medical Image Segmentation by Multi-Frequency in Multi-Scale Attention, Juyeon Nam, Nur Suriza Syazwany, Su Jung Kim, Sang-Chul Lee

Decomposing Disease Descriptions for Enhanced Pathology Detection: A Multi-Aspect Vision-Language Pre-training Framework, Vu Minh Hieu Phan, Yutong Xie, Quanhai Qi, Lingqiao Liu, Liyang Liu, Bowen Zhang, Zhibin Liao, Qi Wu, Minh-Son To, Johan W. Verjans


Fully Convolutional Sliceto-Volume Reconstruction for Single-Stack MRI, Sean I. Young, Oren Kraus, Kian Kenyon-Dean, Saber Saberian, Maryam Falahat, 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, Burton Earnshaw

MLIP: Enhancing Medical Visual Representation with Divergence Encoder and Knowledge-guided Contrastive Learning, Zhe Li, Laurence T. Yang, Bocheng Ren, Xin Nie, Zhangyang Gao, Cheng Tan, Stan Z. Li

FocusMAE: Gallbladder Cancer Detection from Ultrasound Videos with Focused Masked Autoencoders, Soumen Basu, Mayuna Gupta, Chetan Madan, Pankaj Gupta, Chetan Arora

Bi-level Learning of Task-Specific Decoders for Joint Registration and One-Shot Medical Image Segmentation, Xin Fan, Xiaolin Wang, Jiaxin Gao, Jia Wang, Zhongxuan Luo, Risheng Liu

PrPSeg: Universal Proposition Learning for Panoramic Renal Pathology Segmentation, Ruijun Deng, Shuo Li, Can Cui, Tianyuang Yao, Jialin Yue, Juming Xiong, Linying Yu, Yifei Wu, Mengmeng Yin, Yu Wang, Shilin Zhao, Yucheng Tang, Haichun Yang, Yuankai Hou

Versatile Medical Image Segmentation Learned from Multi-Source Datasets via Model Self-Disambiguation, Xiaoyang Chen, Hao Zheng, Yuemeng Li, Yuncong Gao, Li Wang, Heping Li, Yong Fan

Masked Autoencoders for Microscopy are Scalable Learners of Cellular Biology, Oren Kraus, Kian Kenyon-Dean, Saber Saberian, Maryam Falahat, 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, Burton Earnshaw

EMC10: Efficient Multi-scale Convolutional Decoding for Medical Image Segmentation, Md Mostafijur Rahman, Mustafa Munir, Radu Marculescu

Neural Underwater Scene Representation, Yunkai Tang, Chengxuan Zhu, Renjie Wang, Chao Xu, Boxin Shi
PROGRAM GUIDE

Thursday, June 20

208 Hearing Anything Anywhere, Mason Long Wang, Ryosuke Sawata, Samuel Clarke, Ruohan Gao, Shangzhe Wu, Jiajun Wu

209 VMINer: Versatile Multi-view Inverse Rendering with Near- and Far-field Light Sources, Fan Fei, Jiayun Tang, Fang Tan, Boxin Shi

210 EventPS: Real-Time Photometric Stereo Using an Event Camera, Bohan Yu, Jieji Ren, Jin Han, Feishi Wang, Jinxiu Liang, Boxin Shi

211 DiLiGenRT: A Photometric Stereo Dataset with Quantified Roughness and Translucency, Heng Guo, Jieji Ren, Feishi Wang, Boxin Shi, Mingjun Ren, Yasuyuki Matsushita

212 NeRSP: Neural 3D Reconstruction for Reflective Objects with Sparse Polarized Images, Yufei Han, Heng Guo, Koki Fukui, Hiroaki Santo, Boxin Shi, Fumio Okura, Zhanui Ma, Yunpeng Jia

213 EvDIG: Event-guided Direct and Global Components Separation, Xinyu Zhou, Peiqi Duan, Boyu Li, Chu Zhou, Chao Xu, Boxin Shi

214 Differentiable Display Photometric Stereo, Seokjun Choi, Seungwoon Yoon, Giljoo Nam, Seungyong Lee, Seung-Hwan Baek

215 Bayesian Differentiable Physics for Cloth Digitalization, Deshan Gong, Ningtao Mao, He Wang

216 Atlantis: Enabling Underwater Depth Estimation with Stable Diffusion, Fan Zhang, Shaodi You, Yu Li, Ying Fu

217 Sparse Views Near Light: A Practical Paradigm for Uncalibrated Point-light Photometric Stereo, Mohammad Brahimi, Bjorn Haefner, Zhenzhang Ye, Bastian Golduelvec, Daniel Cremers

218 Diffusion Reflectance Map: Single-Image Stochastic Inverse Rendering of Illumination and Reflectance, Yuto Enyo, Ko Nishino


220 Physics-guided Shape-from-Template: Monocular Video Perception through Neural Surrogate Models, David Stotko, Nils Wandell, Reinhardt Klein

221 Spin-UP: Spin Light for Natural Light Uncalibrated Photometric Stereo, Zongrui Li, Zhan Lu, Haojie Yan, Boxin Shi, Gang Pan, Qian Zhang, Xudong Jiang

222 Discontinuity-preserving Normal Integration with Auxiliary Edges, Hyomin Kim, Yuechel Jung, Seungyong Lee

223 A Theory of Joint Light and Heat Transport for Lambertian Scenes, Mani Ramanagopal, Srimat Narayanan, Aswin C. Sankaranarayanan, Sarvina G. Narasimhan

224 IDGuard: Robust General Identity-centric POI Proactive Defense Against Face Editing Abuse, Yunsu Dai, Jianwei Fei, Fangjun Huang

225 Ungeneralizable Examples, Jingwen Ye, Xinchao Wang

226 Distilled Datamodel with Reverse Gradient Matching, Jingwen Ye, Ruanan Yu, Songhua Liu, Xinchao Wang

227 EditGuard: Versatile Image Watermarking for Tamper Localization and Copyright Protection, Xuanyu Zhang, Runyi Li, Jiwen Yu, Youmin Xu, Weiqi Li, Jian Zhang

228 SocialCounterfactuals: Probing and Mitigating Intersectional Social Biases in Vision-Language Models with Counterfactual Examples, Phillip Howard, Avinash Madaus, Tiop Le, Gustavo Lujan Moreno, Anahita Bhawandwala, Vasudev Lal

229 FedAS: Bridging Inconsistency in Personalized Federated Learning, Xiyuan Yang, Wenke Huang, Yang Ye

230 FairRAG: Fair Human Generation via Fair Retrieval Augmentation, Robik Shrestha, Yang Zou, Qiuyu Chen, Zhiheng Li, Yusheng Xie, Siqi Deng

231 Self-Discovering Interpretable Depth Latent Directions for Responsible Text-to-Image Generation, Hang Li, Chengzhi Shen, Philip Torr, Volker Tresp, Jindong Gu

232 ExMap: Leveraging Explainability Heatmaps for Unsupervised Group Robustness to Spurious Correlations, Rwidddhi Chakraborty, Adrian Stetten, Michael C. Kampffmeyer

233 Data Valuation and Detections in Federated Learning, Wengian Li, Shuran Fu, Fengyu Zhang, Yang Pan

234 Utility-Fairness Trade-Offs and How to Find Them, Sepehr Dehdashtian, Bashir Sadeghi, Vishnu Naresh Boddeti

235 SimAC: A Simple Anti-Customization Method for Protecting Face Privacy against Text-to-Image Synthesis of Diffusion Models, Feifei Wang, Zhentao Tan, Tianyi Wei, Yue Wu, Qidong Huang

236 GLOW: Global Layout Aware Attacks on Object Detection, Jun Bao, Buyu Liu, Kui Ren, Jun Yu

237 FADES: Fair Disentangled Representation with Sensitive Relevance, Taekjip Jang, Xiaohang Wang

238 Fair Federated Learning under Domain Skew with Local Consistency and Domain Diversity, Yuhang Chen, Wenke Huang, Mang Ye

239 WaterRF: Robust Watermarks in Radiance Fields for Protection of Copyrights, Youngdong Jang, Dong In Lee, MinHyuk Jang, Jong Wook Kim, Feng Yang, Sangil Kim

240 FLHetBench: Benchmarking Device and State Heterogeneity in Federated Learning, Junyuan Zhang, Shuang Zeng, Miao Zhang, Runxi Wang, Feifei Wang, Yuyin Zhou, Paul Pu Liang, Liangqiong Qu

241 An Upload-Efficient Scheme for Transferring Knowledge From a Server-Side Pre-trained Generator to Clients in Heterogeneous Federated Learning, Jiangqing Zhang, Yang Liu, Yang Hua, Jian Cao

242 Privacy-Preserving Options for Enhancing Protection in Face De-Identification, Jhon Lopez, Carlos Hinojosa, Henry Arguello, Bernard Ghanem

243 A Stealthy Wrongdoer: Feature-Oriented Reconstruction Attack against Split Learning, Xiaoyang Xu, Mengda Yang, Wenzhe Yi, Ziang Li, Juan Wang, Hongxin Yu, Yong Zhang, Yaxin Liu

244 RCL: Reliable Continual Learning for Unified Failure Detection, Fei Zu, Zhen Cheng, Xu-Yao Zhang, Cheng-Lin Liu, Zhaoxiang Zhang

245 Global and Local Prompts Cooperation via Optimal Transport for Federated Learning, Hongxia Li, Wei Huang, Jingya Wang, Ye Shi


247 Explaining CLIP’s Performance Disparities on Data from Blind/Low Vision Users, Daniela Masaceti, Camilla Longden, Agnieszka Slowik, Samuel Wills, Martin Grayson, Cecily Morrison

248 Model Inversion Robustness: Can Transfer Learning Help?, Ky-Tuyen Ho, Koh Jun Hoo, Keshigeyan Chandrasegaran, Ngoc-Bao Nguyen, Ngai-Mai Cheung

249 Make Me a BNN: A Simple Strategy for Estimating Bayesian Uncertainty from Pre-trained Models, Gianni Franchi, Olivier Laurent, Maexecne Legueru, Andrei Bursuc, Andrea Pilzer, Angela Yao

250 Validating Privacy-Preserving Face Recognition under a Minimum Assumption, Hui Zhang, Xingbo Dong, YanLung Lai, Ying Zhou, Xiaoyan Zhang, Xinguo Lu, Zhe Jin, Xuejun Li

251 Re-thinking Data Availability Attacks Against Deep Neural Networks, Bin Fang, Bo Li, Shuang Wu, Shouhong Ding, Ran Yi, Lizhuang Ma

252 OpenBias: Open-set Bias Detection in Text-to-Image Generative Models, Moreno D’Incà, Elia Peruzzo, Massimiliano Mancini, Deja Xu, Vidit Goel, Xiaoyang Xu, Ziyang Wang, Humphrey Shi, Nicu Sebe

253 In-distribution Public Data Synthesis with Diffusion Models for Differentially Private Image Classification, Jinseong Park, Yujin Choi, Jaewook Lee

254 Leak and Learn: An Attacker’s Cookbook to Train Using Leaked Data from Federated Learning, Joshua C. Zhao, Ahaan Dabholkar, Atul Sharma, Saurabh Bagchi

255 Countering Personalized Text-to-Image Generation with Influence Watermarks, Hanwen Liu, Zhicheng Sun, Yadong Mu

256 Fair-VPT: Fair Visual Prompt Tuning for Image Classification, Sungho Park, Hyeran Byun

257 Relaxed Contrastive Learning for Federated Learning, Seonguk Seo, Jinkyu Kim, Geeho Kim, Bohyung Han

258 FairCLIP: Harnessing Fairness in Vision-Language Learning, Yan Luo, Min Shi, Muhammad Osama Khan, Muhammad Muneeb Afzal, Hao Huang, Shuaihang Yuan, Yu Tian, Luo Song, Ava Kourhan, Tobias Elze, Yi Fang, Mengyu Wang

259 Steganographic Passport: An Owner and User Verifiable Credential for Deep Model IP Protection Without Retraining, Qi Cui, Ruohan Meng, Chaohui Xu, Chip-Hong Chang

260 Adaptive Hyper-graph Aggregation for Modality-Agnostic Federated Learning, Fan Qi, Shuai Li
Navigate Beyond Shortcuts: Debiased Learning Through the Lens of Neural Collapse, Yining Wang, Junjie Sun, Chenyue Wang, Mi Zhang, Min Yang

Enhancing Intrinsic Features for Debiasing via Investigating Class-Discerning Common Attributes in Bias-Contrastive Pair, Jeonghoon Park, Chaeyeon Chang, Jaegul Choo

Device-Wise Federated Network Pruning, Shangqian Gao, Junyi Li, Zeyu Zhang, Yanfu Zhang, Weidong Cai, Heng Huang

All Rivers Run to the Sea: Private Learning with Asymmetric Flows, Yue Niu, Ramy E. Ali, Saurav Prakash, Salman Avestimehr

VA3: Virtually Assured Amplification Attack on Probabilistic Copyright Protection for Text-to-Image Generative Models, Xiang Li, Qianli Shen, Kenji Kawaguchi

CPR: Retrieval Augmented Generation for Copyright Protection, Aditya Gotalkar, Alessandro Achille, Luca Zancato, Yu-Xiang Wang, Ashwin Swaminathan, Stefano Soatto

Communication-Efficient Federated Learning with Accelerated Client Gradient, Gheo Kim, Jinkyu Kim, Botihung Han

Self-supervised Debiasing Using Low Rank Regularization, Geon Yeong Park, Chanhyong Jung, Sangmin Lee, Jong Chul Ye, Sang Wan Lee

Facial Identity Anonymization via Intrinsic and Extrinsic Attention Distraction, Zhenzhong Kuang, Xiaochen Yang, Yingjie Shen, Chao Hu, Jun Yu

Collaborative Learning of Anomalies with Privacy (CLAP) for Unsupervised Video Anomaly Detection: A New Baseline, Anas Al-Ijaham, Muhammad Zaigham Zaheer, Nurbek Tastan, Karrthik NANDAKUMAR

Label-Efficient Group Robustness via Out-of-Distribution Concept Curation, Yiwei Yang, Anthony Z. Liu, Robert Wolfe, Aylin Caliskan, Bill Howe

Long-Tailed Anomaly Detection with Learnable Class Names, Chih-Hui Ho, Kuan-Chuan Peng, Nuno Vasconcelos

Robust Emotion Recognition in Context Debiasing, Dingkang Yang, Kun Yang, Mingcheng Li, Shunli Wang, Shuaibing Wang, Lihua Zhang

Correlation-Decoupled Knowledge Distillation for Multimodal Sentiment Analysis with Incomplete Modalities, Mingcheng Li, Dingkang Yang, Xiao Shaohuan, Shuaibing Wang, Yan Wang, Kun Yang, Mingyang Sun, Dongliang Kou, Ziyun Qian, Lihua Zhang

An Edit Friendly DDPM Noise Space: Inversion and Manipulations, Inbar Huberman-Spiegelglas, Vladimir Kulikov, Tamer Michaeli

SleepVST: Sleep Staging from Near-Infrared Video Signals using An Edit Friendly DDPM Noise Space, Inbar Huberman-Spiegelglas, Vladimir Kulikov, Tamer Michaeli

Pre-Trained Transformers, Jonathan F. Carter, Joao Jorge, Oliver Gibson, Lionel Tarassenko

AM-RADIO: Agglomerative Feature Formation Model Reduce All Domains Into One, Mike Ran zig, Greg Heinrich, Jan Kautz, Pavlo Molchanov

Towards Language-Driven Video Inpainting via Multimodal Large Language Models, Jianzong Wu, Xiangtai Li, Chensiyang Si, Shangchen Zhou, Jungkang Yang, Jiangning Zhang, Yining Li, Kai Chen, Yunhui Tang, Ziwei Liu, Chen Change Loy

FedSOL: Stabilized Orthogonal Learning with Proximal Restrictions in Federated Learning, Gihun Lee, Minchan Jeong, Sangmook Kim, Jaehoon Oh, Se-Young Yun

UnionFormer: Unified-Learning Transformer with Multi-View Representation for Image Manipulation Detection and Localization, Shuaibo Li, Wei Ma, Jianwei Guo, Shihiao Xu, Benchong Li, Xiaoqiang Zhang

Motion Blur Decomposition with Cross-shutter Guidance, Xiang Ji, Haiyang Jiang, Yingqiang Zheng


Rapid 3D Model Generation with Intuitive 3D Input, Tianrun Chen, Chaotao Ding, Shangzheng Zhang, Chunan Yu, Ying Zang, Zejian Li, Sida Peng, Linyun Sun

SketchINR: A First Look into Sketches as Implicit Neural Representations, Hmrishav Bandyopadhyay, Ayan Kumar Bhunia, Pinaki Nath Chowdhury, Aneeshen Sain, Tao Xiang, Timothy Hospedales, Yi-Zhe Song


DiaLoc: An Iterative Approach to Embodied Dialog Localization, Chao Zhang, Mohan Li, Ignas Budvytis, Stephan Liwicki

WildlifeMapper: Aerial Image Analysis for Multi-Species Detection and Identification, Satish Kumar, Bowen Zhang, Chandrakant Gudavalli, Connor Levenson, Lacey Hughey, Jared A. Stabach, Irene Amoke, Gordon Ojwang, Joseph Mukeka, Stephen Mwiu, Joseph Ogutu, Howard Frederick, B.S. Manjunath

Harnessing Meta-Learning for Improving Full-Frame Video Stabilization, Muhammad Kashif Ali, Eun Woo Im, Dongjin Kim, Tae Hyun Kim

De-confounded Data-free Knowledge Distillation for Handling Distribution Shifts, Yuzheng Wang, Dingkang Yang, Zhouyu Chen, Yang Liu, Siao Liu, Weniqiang Zhang, Liuhua Zhang, Lizhe Qi

Day-Night Cross-domain Vehicle Re-identification, Hongchao Li, Jingong Chen, Aihua Zhang, Yong Wu, Yonglong Luo

Brush2Prompt: Contextual Prompt Generator for Object Inpainting, Mang Tik Chiu, Yuqian Zhou, Lingzhi Zhang, Zhe Lin, Connally Nesbitt, Sohrab Amighiodsi, Eli Shechtman, Humphrey Shi

Cloud-Device Collaborative Learning for Multimodal Large Language Models, Guangjun Wang, Jianming Liu, Chenxuan Liu, Yuan Zhang, Junpeng Ma, Xinyu Wei, Kevin Zhang, Maurice Chong, Rerun Zhang, Yijiang Liu, Shanghang Zhang

Making Visual Sense of Oracle Bones for You and Me, Runqi Qiao, Lan Yang, Kuye Pang, Honggang Zhang

Boosting Object Detection with Zero-Shot Day-Night Domain Adaptation, Zhipeng Du, Miaojing Shi, Jiankang Deng

InNeRF360: Text-Guided 3D-Consistent Object Inpainting on 360-degree Neural Radiance Fields, Dongqing Wang, Tong Zhang, Alaa Abboud, Sabine Süsstrunk

Language Models as Black-Box Optimizers for Vision-Language Models, Shihong Liu, Samuel Yu, Zhiqiu Lin, Deepak Pathak, Deva Ramanan

Mind Marginal Non-Crack Regions: Clustering-Inspired Representation Learning for Crack Segmentation, Zhuzhuanghui Chen, Zhuoan Lai, Jie Chen, Jianqiang Li

InstructDiffusion: A Generalist Modeling Interface for Vision Tasks, Ziqing Geng, Binxin Yang, Tiankai Hang, Chen Li, Shuyang Gu, Ting Zhang, Jianmin Bao, Zheng Zhang, Houqiang Li, Han Hu, Dong Chen, Baining Guo


Physical Backdoor: Towards Temperature-based Backdoor Attacks in the Physical World, Wen Yin, Jian Lou, Pan Zhou, Yulai Xie, Dan Feng, Yuhua Sun, Tailai Zhang, Lichao Sun

Behind the Veil: Enhanced Indoor 3D Scene Reconstruction with Occluded Surfaces Completion, Su Sun, Cheng Zhao, Yuliang Guo, Ruoyu Wang, Xinyu Huang, Yingjie Victor Chen, Liu Ren

EarthLoc: Astronaut Photography Localization by Indexing Earth from Space, Gabrielle Bertol, Alex Stoken, Barbara Caputo, Carlo Masone

DiffForeskins: Leveraging Diffusion Prior to Image Forgery Detection and Localization, Zeqin Yu, Jingquan Ni, Yuzhen Lin, Haoyi Deng, Bin Li


The Unreasonable Effectiveness of Pre-Trained Features for Camera Pose Refinement, Gabriele Trivigno, Carlo Masone, Barbara Caputo, Torsten Sattler

Cloud Image Quality Assessment Based on Geometric Order Learning, Nyeong-Ho Shin, Seon-Ho Lee, Chang-Su Kim

CrowdDiff: Multi-hypothesis Crowd Density Estimation using Diffusion Models, Yasiru Ranasinghe, Nithin Gopalkrishnan Nair, Wele Gedara Chaminda Bandara, Vishal M. Patel

Towards Efficient Reply in Federated Incremental Learning, Yichen Li, Quanwei Li, Haozhao Wang, Ruixuan Li, Wenliang Zhang, Guannan Zhang
309 MART: Masked Affective RepresenTation Learning via Masked Temporal Distribution Distillation, Zhicheng Zhang, Pancheng Zhao, Eunil Park, Jufeng Yang
310 PolarRec: Improving Radio Interferometric Data Reconstruction Using Polar Coordinates, Ruoci Wang, Zhuoyuan Chen, Jiayi Zhu, Qiong Luo, Feng Wang
311 Constrained Layout Generation with Factor Graphs, Mohammed Haroon Dupaty, Yanfei Dong, Sicleng Leng, Guojia Fu, Yong Liang Goh, Wei Lu, Wee Sun Lee
312 Visual In-Context Prompting, Feng Li, Qing Jiang, Hao Zhang, Tianhe Ren, Shilong Liu, Xueyan Zou, Huaihue Xu, Hongyang Li, Jianwei Yang, Chunnian Li, Lei Zhang, Jianfeng Gao
313 Traceable Federated Continual Learning, Qiang Wang, Bingyan Liu, Yawen Li
314 Interactive Continual Learning: Fast and Slow Thinking, Biqing Qi, Xinjuan Chen, Junqi Gao, Dong Li, Xianing Liu, Ligang Wu, Bowen Zhou
315 PIGEON: Predicting Image Geocodes,

316 Lukas Haas, Michael Skreta, Silas Alberti, Chelsea Finn
317 LQMFormer: Language-aware Query Mask Transformer for Referencing Image Segmentation, Nisarg A. Shah, Vibashan VS, Vishal M. Patel
318 VIP-LLaVA: Making Large Multimodal Models Understand Arbitrary Visual Prompts, Mu Cai, Haotian Liu, Siva Karthik Mustikovela, Gregory P. Meyer, Yuning Chai, Dennis Park, Yong Jae Lee
319 DePT: Decoupled Prompt Tuning, Ji Zhang, Shihan Wu, Lianli Gao, Heng Tao Shen, Jingkuan Song
320 Grounded Question-Answering in Long Epocentric Videos, Shangzhe Di, Weidi Xie
321 HalluciDoctor: Mitigating Hallucinatory Toxicity in Visual Instruction Data, Qifan Yu, Juncheng Li, Longhui Wei, Liang Pang, Wentao Ye, Bosheng Qin, Silin Tang, Qi Tian, Yueting Zhuang
322 ViTamin: Designing Scalable Vision Models in the Vision-Language Era, Jienneng Chen, Qihang Yu, Xiaohui Shen, Alan Yuille, Liang-Chieh Chen
323 The Manga Whisperer: Automatically Generating Transcriptions for Comics, Ragav Sachdeva, Andrew Zisserman
324 Learning to Localize Objects Improves Spatial Reasoning in Visual-LLMs, Kanchana Ranasinghe, Satya Narayan Shukla, Omid Poursaeed, Michael S. Ryoo, Tsung-Yu Lin
325 The Neglected Tails in Vision-Language Models, Shubham Parashar, Zhiqiu Lin, Tian Liu, Xiangjue Dong, Yanan Li, Deva Ramanan, James Caverlee, Shuo Kong
326 Unveiling Parts Beyond Objects: Towards Finer-Granularity Referring Expression Segmentation, Wenxuan Wang, Tongtian Yue, Yisi Zhang, Longteng Guo, Xingjian He, Xinglong Wang, Jing Liu
327 GLaMM: Pixel Grounding Large Multimodal Model, Hanooona Rasheed, Muhammad Maaz, Sahal Shahyi, Abdelrahman Shaker, Salman Khan, Hisham Chehak, Rao M. Anwer, Eric Xing, Ming-Hsuan Yang, Fahad S. Khan
328 Alpha-CLIP: A CLIP Model Focusing on Wherever You Want, Zeyi Sun, Ye Fang, Tong Wu, Pan Zhang, Yuhang Zang, Shuo Kong, Yuanjun Xiong, Dahua Lin, Shiyuan Wei, Tongtian Yue, Jie Cheng, Wenxuan Wang, Tongtian Yue, Jie Cheng
329 SC-Tune: Unleashing Self-Consistent Referential Comprehension in Large Vision Language Models, Tongtian Yue, Jie Cheng, Longteng Guo, Xingyuan Dai, Zijia Zhao, Xingjian He, Gang Xiong, Yisheng Lv, Jing Liu
330 V?: Guided Visual Search as a Core Mechanism in Multimodal LLMs, Penghao Wu, Saining Xie
331 Improved Visual Grounding through Self-Consistent Explanations, Ruozhen He, Paola Cascante-Bonilla, Ziyan Yang, Alexander C. Berg, Vicente Ordonez
332 Distilling Vision-Language Models on Millions of Videos, Yue Zhao, Long Zhao, Xingyu Zhou, Jialin Wu, Chun-Te Chu, Hui Miao, Florian Schroff, Hartwig Adam, Ting Liu, Boqing Gong, Philipp Krahenbuhl, Liangzhe Yuan
333 Separating the “Chirp” from the “Chat”: Self-supervised Visual Grounding of Sound and Language, Mark Hamilton, Andrew Zisserman, John R. Hershey, William T. Freeman
334 Referring Image Editing: Object-level Image Editing via Referring Expressions, Chang Liu, Xiangtai Li, Henghui Ding
335 Vision-and-Language Navigation via Causal Learning, Liuyi Wang, Zongtao He, Ronghao Dang, Mengjiao Shen, Chengjiu Liu, Qijun Chen
336 VISTA-LLAMA: Reducing Hallucination in Video Language Models via Equal Distance to Visual Tokens, Fan Ma, Xiaojie Jin, Heng Wang, Yuchen Xian, Jiashii Feng, Yi Yang
337 Ranking Distillation for Open-Ended Video Question Answering with Insufficient Labels, Tiansheng Liang, Chaolei Tan, Beihao Xia, Wei-Shi Zheng, Jian-Fang Hu
338 Can I Trust Your Answer? Visually Grounded Video Question Answering, Junbin Xiao, Angela Yao, Yicong Li, Tat-Seng Chua
339 Prompt Highlighter: Interactive Control for Multi-Modal LLMs, Yuechen Zhang, Shengyu Qian, Bohaoh Feng, Shu Liu, Jiaya Jia
340 Language-only Training of Zero-shot Composed Image Retrieval, Geonmu Gu, Sanghyuk Chun, Wonje Kim, Yoohoon Kang, Sangdoo Yun
341 MoReVQA: Exploring Modular Reasoning Models for Video Question Answering, Juhong Min, Shyamal Buch, Arsha Nagrani, Minsu Cho, Cordelia Schmid
342 Let’s Think Outside the Box: Exploring Leap-of-Thought in Large Language Models with Creative Humor Generation, Shanshan Zhong, Zhongzhan Huang, Shanghua Gao, Wushao Wen, Liang Lin, Marinka Zitnik, Pan Zhou
343 CLOVA: A Closed-Loop Visual Assistant with Tool Usage and Update, Zhi Gao, Yuntao Du, Xiong Zhang, Xiaojian Ma, Wenjuan Han, Song-Chun Zhu, Qing Li
344 Language-only Training of Zero-shot Composed Image Retrieval, Geonmu Gu, Sanghyuk Chun, Wonje Kim, Yoohoon Kang, Sangdoo Yun
345 Synthesizes Diagnose and Optimize: Towards Fine-Grained Vision-Language Understanding, Wujian Peng, Sicheng Xie, Zuyao You, Shiyi Lan, Zuxuan Wu
346 AssistGUI: Task-Oriented PC Graphical User Interface Automation, Difei Gao, Lei Ji, Zechen Bai, Minyong Ouyang, Peiran Li, Dongxing Mao, Qinchen Wu, Weichen Zhang, Peiyi Wang, Xiangyu Guo, Hengxu Wang, Luoweii Zhou, Mike Zheng Shou
347 SEED-Bench: Benchmarking Multimodal Large Language Models, Bohao Li, Youying Ge, Yixiao Ge, Guangzhi Wang, Rui Wang, Ruimao Zhang, Ying Shan
348 Unknown Prompt the only Lacuna: Unveiling CLIP’s Potential for Open Domain Generalization, Mainak Singha, Ankit Jha, Shirshe Bose, Ashwin Nair, Miloud Abbad, Biplab Banerjee
349 Panda-70M: Captioning 70M Videos with Multiple Cross-Modality Teachers, Tais-Shien Chen, Alaksandr Siarohin, Willi Menapace, Ekaterina Deyneka, Hsiang-wei Chao, Byung Eun Jeon, Yuewei Fang, Hsin-Ying Lee, Jian Ren, Ming-Hsuan Yang, Soroush Tulyakov
350 Decoupling Static and Hierarchical Motion Perception for Referring Video Segmentation, Shuting He, Henghui Ding
351 Causal-CoG: A Causal-Effect Look at Context Generation for Boosting Multi-modal Language Models, Shitian Zhao, Zhiowuan Li, Yadong Lu, Alan Yuille, Yan Wang

PROGRAM GUIDE

THURSDAY, JUNE 20

34 | CVPR 2024 | PROGRAM GUIDE MAIN CONFERENCE
10:30 - 18:45 Art Program (Arch 4CDE)

11:00 - 13:00 Orals 4A: Autonomous Navigation and Egocentric Vision (Summit Elliott Bay Room)

12:00 - 14:00 LUNCH (Summit ExHall 1-2)

12:00 - 13:00 SOCIAL - Student Speed Mentorship Session (Summit Elliott Bay Room)

**Reservation Required**

13:00 - 14:30 Orals 4A: Autonomous Navigation and Egocentric Vision (Summit Ballroom)

1. SAFDNet: A Simple and Effective Network for Fully Sparse 3D Object Detection, Gang Zhang, Junnan Chen, Guohuan Gao, Jianmin Li, Si Liu, Xiaolin Hu

2. UnO: Unsupervised Occupancy Fields for Perception and Forecasting, Ben Agro, Quinlan Sykora, Sergio Casas, Thomas Gilles, Raquel Urtasun
3. **EgoGen: An Egocentric Synthetic Data Generator**, Gen Li, Kaiyue Cao, Siwei Zhang, Xiaohong Lyu, Mihai Dusmanu, Yan Zhang, Marc Pollefeys, Yu Tu

4. **Learning to Segment Referenced Objects from Narrated Egocentric Videos**, Yuhan Shen, Huiyu Wang, Xiting Yang, Matt Feiszli, Ehsan Elhamifar, Lorenzo Torresani, Effrosyni Mavraki

5. Producing and Leveraging Online Map Uncertainty in Trajectory Prediction, Junxiang Gu, Guanyu Song, Igor Gilitschenski, Marco Pavone, Boris Ivanovic

### 13:00 - 14:30 **Orals 4B: 3D Vision (Summit Flex Hall AB)**

1. **SceneFun3D: Fine-Grained Functionality and Affordance Understanding in 3D Scenes**, Alex Deltizas, Ayca Tekmaz, Federico Tombari, Roberto Sumner, Marc Pollefeys, Francis Engelmann

2. **SpiderMatch: 3D Shape Matching with Global Optimality and Geometric Consistency**, Paul Roettger, Florian Bernard

3. **PaScO: Urban 3D Panoptic Scene Completion with Uncertainty Awareness**, Anh-Quan Cao, Angela Dai, Raoul de Charette

4. **PlatoNeRF: 3D Reconstruction in Plato's Cave via Single-View Two-Bounce Lidar**, Tzofi Klinghoffer, Xiaoyu Xiang, Siddharth Somasundaram, Yuchen Fan, Christian Richardt, Ramesh Raskar, Rakesh Ranjan

5. A Subspace-Constrained Tyler’s Estimator and its Applications to Structure from Motion, Feng Yu, Teng Zhang, Gilad Lerman

### 13:00 - 14:30 **Orals 4C: Action and Motion (Summit Flex Hall C)**


2. An N-Point Linear Solver for Line and Motion Estimation with Event Cameras, Ling Gao, Daniel Gehrig, Hang Su, Davide Scaramuzza, Laurent Kneip

3. **RoHM: Robust Human Motion Reconstruction via Diffusion**, Siwei Zhang, Bharat Lal Bhattacharjee, Yuanlu Xu, Alexander Winkler, Petar Kadlecsek, Siyu Tang, Federica Bogo

4. Temporally Consistent Unbalanced Optimal Transport for Unsupervised Action Segmentation, Ming Xu, Stephen Gould


### 14:30 - 14:45 **Courtesey Break**

### 14:45 - 15:45 **KEYNOTE 2 - David Baker, Director of Institute for Protein Design**

- **Hennieta and Aubrey Davis Endowed Professor in Biochemistry**, University of Washington (Summit Flex Hall ABC)

### 15:45 - 16:00 **Courtesey Break**

### 16:00 - 17:00 **PAMI TC Meeting (Summit Flex Hall ABC)**

### 16:45 - 17:15 **Poster Setup**

### 17:00 - 19:00 **SOCIAL - CV Entrepreneurship — Founders, Freelancers & Friends**

### 17:15 - 18:45 **Poster Session 4 & Exhibit Hall**

1. **PaScO: Urban 3D Panoptic Scene Completion with Uncertainty Awareness**, Anh-Quan Cao, Angela Dai, Raoul de Charette

2. **PlatoNeRF: 3D Reconstruction in Plato’s Cave via Single-View Two-Bounce Lidar**, Tzofi Klinghoffer, Xiaoyu Xiang, Siddharth Somasundaram, Yuchen Fan, Christian Richardt, Ramesh Raskar, Rakesh Ranjan

3. **Self-Supervised Class-Agnostic Motion Prediction with Spatial and Temporal Consistency Regularizations**, Rongjing Wang, Yizheng Xu, Jun Cen, Zhiyu Li, Zhe Wu, Zhangyu Cao, Guosheng Lin

4. **Multi-Space Alignments Towards Universal LiDAR Segmentation**, Youquan Liu, Jingdong Kong, Xiaoyang Wu, Runnan Chen, Xin Li, Liang Pan, Zhiwei Liu, Yuxin Ma

5. **Generalized Predictive Model for Autonomous Driving**, Jiazi Yang, Shenyuan Guo, Yihang Qiu, Li Chen, Tianyu Li, Bo Dai, Kasyapa Chitta, Penghao Wu, Jia Zeng, Ping Luo, Jun Zhang, Andreas Geiger, Yu Qiao, Hongyang Li

6. **Visual Point Cloud Forecasting enables Scalable Autonomous Driving**, Zetong Yang, Li Chen, Yanan Sun, Hongyang Li


8. **AIDE: An Automatic Data Engine for Object Detection in Autonomous Driving**, Mingfu Liang, Jong-Chyi Su, Samuel Schuler, Sparsh Garg, Shiyu Zhao, Ying Wu, Manmohan Chandraker

9. Dynamic Adapter Meets Prompt Tuning: Parameter-Efficient Transfer Learning for Point Cloud Analysis, Xin Zhou, Dingkang Liang, Wei Xu, Xingkui Zhu, Yihan Xu, Zhikang Zou, Xiang Bai


12. **Towards Realistic Scene Generation with LiDAR Diffusion Models**, Haoxi Ran, Vitor Guizilini, Yie Wang

13. **Driving into the Future: Multiview Visual Forecasting and Planning with World Model for Autonomous Driving**, Yuqi Wang, Jiawei He, Lue Fan, Hongxin Li, Yuntao Chen, Zhaoxiang Zhang


15. **Scaling Diffusion Models to Real-World 3D LiDAR Scene Completion**, Lucas Nunes, Rodrigo Marcuzzi, Benedikt Mersch, Jens Behley, Cyril Stachniss

16. **UniMix: Towards Domain Adaptive and Generalizable LiDAR Semantic Segmentation in Adverse Weather**, Haimei Zhao, Jing Zhang, Zhuo Chen, Shanshan Zhao, Dacheng Tao

17. **Not All Voxels Are Equal: Hardness-Aware Semantic Scene Completion with Self-Distillation**, Song Wang, Jiawei Yu, Wenyong Li, Wenyu Liu, Xiaolu Liu, Junbo Chen, Jianke Zhu


19. **MGMap: Mask-Guided Learning for Online Vectorized HD Map Construction, Xiaolu Liu, Song Wang, Wentong Li, Ruiyang Li, Junbo Chen, Jianke Zhu**

20. **Density-Adaptive Model Based on Motif Matrix for Multi-Agent Trajectory Prediction**, Di Wen, Haoran Xu, Zhaocheng He, Zhe Wu, Guang Tan, Peixi Peng

21. **StreamingFlow: Streaming Occupancy Forecasting with Asynchronous Multi-modal Data Streams via Neural Ordinary Differential Equation**, Yining Shi, Kun Jiang, Ke Wang, Jiushi Li, Yunlong Wang, Mengmeng Yang, Diangye Yang

22. **SAFDNet: A Simple and Effective Network for Fully Sparse 3D Object Detection**, Gang Zhang, Junnan Chen, Guohuan Gao, Jianmin Li, Si Liu, Xiaolin Hu

23. View From Above: Orthogonal-View aware Cross-view Localization, Shan Wang, Chuong Nguyen, Jiawei Liu, Yanhao Zhang, Sundaram Muthu, Fahira Afzal Maken, Kaibao Zhang, Hongdong Li


25. **Is Ego Status All You Need for Open-Loop End-to-End Autonomous Driving?**, Zhiqi Li, Zhihong Yu, Shiyi Lan, Jiahai Li, Jian Kautz, Tong Lu, Jose M. Alvarez


27. **Adversarial Backdoor Attack by Naturalistic Data Poisoning on Trajectory Prediction in Autonomous Driving**, Mozghan Pourkeshavar, Md. Leen T. Sabokrou, Amir Rassooli


29. **IS-Fusion: Instance-Scene Collaborative Fusion for Multimodal 3D Object Detection**, Junbo Yin, Jianbing Shen, Runnan Chen, Wei Li, Ruigang Yang, Pascal Frossard, Wenguan Wang
30 LSK3DNet: Towards Effective and Efficient 3D Perception with Large Sparse Kernels, Tuo Feng, Wenguang Wang, Fan Ma, Yi Yang
31 RCB3DNet: Radar-camera Fusion in Bird’s Eye View for 3D Object Detection, Zhwei Lin, Zhe Liu, Zhongyu Xia, Xinhao Wang, Yongtao Wang, Shengxiang Qi, Yang Dong, Nan Dong, Le Zhang, Ce Zhu
32 PTT: Point-Projection Transformer for Efficient Temporal 3D Object Detection, Kuan-Chih Huang, Weijie Lyu, Ming-Hsuan Yang, Yi-Hsuan Tsai
33 Driving Everywhere with Large Language Model Policy Adaptation, Boyi Li, Yue Wang, Jiagen Mao, Boris Ivanovic, Sushant Veer, Karen Leung, Marco Pavone
34 Text2Loc: 3D Point Cloud Localization from Natural Language, Yan Xia, Letian Shi, Zifeng Ding, Joao F. Henriques, Daniel Cremers
35 Commonsense Prototype for Outdoor Unsupervised 3D Object Detection, Hai Wu, Shijia Zhao, Xun Huang, Chenglu Wen, Xin Li, Cheng Wang
36 A-Teacher: Asymmetric Network for 3D Semi-Supervised Object Detection, Hanshi Wang, Zhigpeng Zhang, Jin Gao, Weiming Hu
37 MoST: Multi-Modality Scene Tokenization for Motion Prediction, 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
38 Feedback-Guided Autonomous Driving, Jimmyang Yang, Zanning Huang, Arijit Ray, Eshed Ohn-Bar
39 Bootstrapping Autonomous Driving Radars with Self-Supervised Learning, Yiduo Hao, Sohrab Madani, Junfeng Guan, Mohammed Alloolah, Saurabh Gupta, Haitham Hassanieh
40 UnO: Unsupervised Occupancy Fields for Perception and Forecasting, Ben Agro, Quinlan Sykor, Sergio Casas, Thomas Gilles, Raquel Urtasun
41 SIRA: Scalable Inter-frame Relation and Association for Radar Perception, Ryosuke Yata, Pu Wang, Petros Boufounos, Ryuhei Takahashi
42 SparseOcc: Rethinking Sparse Latent Representation for Vision-Based Semantic Occupancy Prediction, Pin Tang, Zhongdao Wang, Guoqing Wang, Jialin Zheng, Xiangjun Ren, Bailan Feng, Chao Ma
43 DiffLoc: Diffusion Model for Outdoor LiDAR Localization, Wen Li, Yupang Yang, Shangshu Yu, Guosheng Hu, Chenglu Wen, Ming Cheng, Cheng Wang
44 Weak-to-Strong 3D Object Detection with X-Ray Distillation, Alexander Gambahidze, Aleksandr Dedukin, Maxim Golyadkin, Maria Razzhivina, Ilya Makarov
45 T4P: Test-Time Training of Trajectory Prediction via Masked Autoencoder and Actor-specific Token Memory, Daehae Park, Jaeseok Jeong, Sung-Hoon Yoon, Jaewoo Jeong, Kuk-Jin Yoon
46 Editable Scene Simulation for Autonomous Driving via Collaborative LLM-Agents, Yuxi Wei, Zi Wang, Yifan Lu, Chenxin Xu, Changxing Liu, Hao Zhao, Silheng Chen, Yanfang Wang
47 Uncertainty-Guided Never-Ending Learning to Drive, Lei Lai, Eshed Ohn-Bar, Sanjay Arora, John Seon Keun Yi
48 On the Road to Portability: Compressing End-to-End Motion Planner for Autonomous Driving, Kaituo Feng, Changsheng Li, Dongchun Ren, Ye Yuan, Guoren Wu
49 DiffFlow3D: Toward Robust Uncertainty-Aware Scene Flow Estimation with Iterative Diffusion-Based Refinement, Jiuming Liu, Guangming Wang, Weicai Ye, Chaokang Jiang, Jinru Han, Zhe Li, Guofeng Zhang, Dalong Du, Hesheng Wang
50 LMDrive: Closed-Loop End-to-End Driving with Large Language Models, Hao Shao, Yuxuan Hu, Letian Wang, Guanglu Song, Steven L. Waslander, Yu Liu, Hongsheng Li
51 SOAC: Spatio-Temporal Overlap-Aware Multi-Sensor Calibration using Neural Radiance Fields, Quentin Herau, Nathan Piasco, Moussab Bennehar, Luis Roldao, Dzmitry Tsishkou, Cyrille Migniot, Pascal Vasseur, Cédric Demonceaux
52 LaMPilot: An Open Benchmark Dataset for Autonomous Driving with Language Model Programs, Yunshe Ma, Can Cui, Xu Cao, Wengiang Ye, Peiran Liu, Juanwu Lu, Amr Abdelraouf, Rohit Gupta, Kyungtae Han, Aniket Bera, James M. Rehg, Ziran Wang
53 GLIDR: Topologically Regularized Graph Generative Network for Sparse LiDAR Point Clouds, Prashant Kumar, Kshitij Madhav Bhat, Vedang Bhupesh Shenvi Nadkarni, Prem Kalra
54 Towards Robust 3D Object Detection with LiDAR and 3D Radar Fusion in Various Weather Conditions, JuYeong Chae, Hyeonseong Kim, Kuk-Jin Yoon
55 3DSLabelling: Boosting 3D Scene Flow Estimation by Pseudo Auto-labeling, Chaokang Jiang, Guangming Wang, Jiuming Liu, Hesheng Wang, Zhuang Ma, Zhenqiang Liu, Zhujin Liang, Yi Shan, Dalong Du
56 ADA-Track: End-to-End Multi-3D Camera 3D Multi-Object Tracking with Alternating Detection and Association, Shuxiao Ding, Lukas Schneider, Marius Cordts, Juergen Gall
57 PointBeV: A Sparse Approach for BeV Predictions, Loick Chambon, Eloi Zablocki, Mickael Chen, Floren Bartocciioni, Patrick Pérez, Matthieu Cord
58 Light the Night: A Multi-Condition Diffusion Framework for Unpaired Low-Light Enhancement in Autonomous Driving, Jinlong Li, Baolu Li, Zhengzhong Tu, Xinyu Liu, Qian Guo, Felix Juefei-Xu, Runsheng Xu, Hongkai Yu
59 CLIP-BEVFormer: Enhancing Multi-View Image-Based BEV Detector with Ground Truth Flow, Chenbin Pan, Burhaneddin Yaman, Senem Velipasalar, Liu Ren
60 Adapting to Length Shift: FlexiLength Network for Trajectory Prediction, Yi Xu, Yun Fu
61 UniPAD: A Universal Pre-training Paradigm for Autonomous Driving, Hongsui Yang, Sha Zhang, Di Huang, Xiaoyang Wu, Haoyi Zhu, Tong He, Shixiang Tang, Hengshuang Zhao, Qibo Qiu, Binbin Lin, Xuefei He, Wanli Ouyang
62 Higher-order Relational Reasoning for Pedestrian Trajectory Prediction, Sungjune Kim, Hyung-gun Chi, Hyerin Lim, Karthik Ramani, Jinkyu Kim, Sangpil Kim
63 HPNet: Dynamic Trajectory Forecasting with Historical Prediction Attention, Xiaolong Tang, Meina Kan, Shiguang Shan, Zhilong Ji, Jinfeng Bai, Xilin Chen
64 LiSA: LiDAR Localization with Semantic Awareness, Bochun Yang, Zijin Li, Wen Li, Zhigpeng Cai, Chenglu Wen, Yu Zhang, Matthias Muller, Cheng Wang
65 SmartRefine: A Scenario-Adaptive Refinement Framework for Efficient Motion Prediction, Yang Zhou, Hao Shao, Letian Wang, Steven L. Waslander, Hongsheng Li, Yu Liu
66 Pseudo Label Refinery for Unsupervised Domain Adaptation on Cross-dataset 3D Object Detection, Zhanwei Zhang, Minhong Chen, Shuai Xiao, Liang Peng, Hengjia Li, Binbin Lin, Ping Li, Wenxiao Wang, Boxi Wu, Deng Cai, Wanli Ouyang
67 Multi-agent Collaborative Perception via Motion-aware Robust Communication Network, Shixin Hong, Yu Liu, Zhi Li, Shaohui Li, You He
68 TASeg: Temporal Aggregation Network for LiDAR Semantic Segmentation, Xiaopei Wu, Yuenan Hou, Xiaoshui Huang, Binbin Lin, Tong He, Xinge Zhu, Yuexin Ma, Boxi Wu, Hailong Liu, Deng Cai, Wanli Ouyang
69 HINTED: Hard Instance Enhanced Detector with Mixed-Density Feature Fusion for Sparsely-Supervised 3D Object Detection, Qiming Xie, Wei Ye, Hai Wu, Shijia Zhao, Leyuan Xing, Xun Huang, Jinhao Deng, Xin Li, Chenglu Wen, Cheng Wang
70 CaKD: Category-aware Knowledge Distillation and Pruning Framework for Lightweight 3D Object Detection, Haonan Zhang, Longjun Liu, Yuqi Huang, Zhao Yang, Xinyu Lei, Bihan Wen, Hesheng Wang, Zhuang Ma, Zhenqiang Liu, Zhujin Liang, Yi Shan, Dalong Du
71 Diffusion-ES: Gradient-free Planning with Diffusion for Autonomous and Instruction-guided Driving, Brian Yang, Hanqyang Su, Nikolaos Gkanatsios, Tsung-Wei Ke, Ayush Jain, Jeff Auton, Marlon Schneider, Marius Cordts, Juergen Gall
72 TULIP: Transformer for Upsampling of LiDAR Point Clouds, Bin Yang, Patrick Pfeundschuh, Roland Siegwart, Marco Hutter, Peyman Moghadam, Vaishakh Patil
73 Bézier Everywhere: All at Once: Learning Drivable Lanes as Bézier Graphs, Hugh Blaney, Hanlin Tian, Hamish Scott, Nils Goldbeck, Chess Stetson, Panagiotis Angeloudis
74 Flow-Guided Online Stereo Rectification for Wide Baseline Stereo, Anush Kumar, Fahim Mannan, Omid Hosseini Jafari, Shile Li, Felix Heide
98 An Empirical Study of Scaling Law for Scene Text Recognition, Miao Rang, Zhenhui Bi, Chuanjian Liu, Yunhe Wang, Kai Han

99 LayoutLLM: Layout Instruction Tuning with Large Language Models for Document Understanding, Choewei Luo, Yufan Shen, Zhaoqin Zhong, Zhi Yu, Cong Yao

100 OmniParser: A Unified Framework for Text Spotting Key Information Extraction and Table Recognition, Jianqiang Wan, Sibo Song, Wenwen Yu, Yuliang Liu, Wenqing Cheng, Fei Huang, Xiang Bai, Cong Yao, Zhibo Yang


102 LayoutFormer: Hierarchical Text Detection Towards Scene Text Understanding, Min Liang, Jia-Wei Ma, Xiaobin Zhu, Jingyan Qin, Xu-Cheng Yin

103 Generating Handwritten Mathematical Expressions From Symbol Graphs: An End-to-End Pipeline, Yu Chen, Fei Gao, Yanguang Zhang, Maoying Qiao, Nannan Wang

104 OpenESS: Event-based Semantic Scene Understanding with Open Vocabularies, Longdong Kong, Youquan Liu, Lai Xing Ng, Benoit R. Cottereau, Wei Tsang Ooi

105 PELA: Learning Parameter-Efficient Models with Low-Rank Approximation, Yangyang Guo, Guangzi Wang, Mohan Kankanhalli

106 MADTP: Multimodal Alignment-Guided Dynamic Token Pruning for Accelerating Vision-Language Transformer, Jianjian Cao, Peng Ye, Shengze Li, Chong Yu, Yansong Tang, Jiwen Lu, Tao Chen

107 VdK: Improving Knowledge Distillation using Orthogonal Projections, Roy Miles, Ismail Elezi, Jiankang Deng

108 Logit Standardization in Knowledge Distillation, Hongtao Sun, Wenyi Ren, Jingzhi Li, Rui Wang, Xiaochun Cao

109 Multi-criteria Token Fusion with One-step-ahead Attention for Efficient Vision Transformers, Sanghyeok Lee, Joonmyung Choi, Hyunwoo J. Kim

110 ParameterNet: Parameters Are All You Need for Large-scale Visual Pretraining of Mobile Networks, Kai Han, Yunhe Wang, Jianyuan Guo, Enhua Wu

111 DeepCache: Accelerating Distillation Models for Free, Xinyi Ma, Gongfang Fan, Xinchao Wang

112 ALGM: Adaptive Local-then-Global Token Merging for Efficient Semantic Segmentation with Plain Vision Transformers, Nargis Norouzi, Svetlana Orlova, Daan de Geus, Gijs Dubbelman

113 A General and Efficient Training for Transformer via Token Expansion, Wenxuan Huang, Yuncheng Zhong, Gaoqi He, Ke Li, Xing Sun, Shaohui Lin

114 Efficient Dataset Distillation via Minimax Diffusion, Jianyang Gu, Saeed Vahidian, Vyacheslav Kurgurtsev, Haonan Wang, Wei Jiang, Yang You, Yiran Chen

115 PEM: Prototype-based Efficient MaskFormer for Image Segmentation, Niccolò Cavagnero, Gabriele Rosi, Claudia Cuttano, Francesca Pittilli, Marco Cicone, Giuseppe Avera, Fabio Carmelli

116 Transferable and Principled Efficiency for Open-Vocabulary Segmentation, Jingxuan Xu, Wuyang Chen, Yao Zhao, YUNCHAO WEI

117 Dense Vision Transformer Compression with Few Samples, Hanxiao Zhang, Yifan Guo, Guo-Hua Wang

118 Dr2Net: Dynamic Reversible Dual-Residual Networks for Memory-Efficient Finetuning, Chen Zhao, Shuming Liu, Karttikeya Mangalam, Guocheng Qian, Fatimah Zohra, Abdulkhaleg Alghannam, Jitendra Malik, Bernard Ghanem

119 MaxQ: Multi-Axis Query for N:M Sparsity Network, Jingyang Xiang, Siqi Li, Junhao Chen, Zhuangzhi Chen, Tianxin Huang, Linpeng Peng, Yong Liu

120 Retraining-Free Model Quantization via One-Shot Weight-Coupling Learning, Chen Tang, Yuan Meng, Jiacheng Jiang, Shuzhao Xie, Rongwei Lu, Xinhua Ma, Zhi Wang, Wenwu Zhu

121 LORS: Low-rank Residual Structure for Parameter-Efficient Network Stacking, iain Li, Qiang Nie, Weifu Fu, Yuhuan Lin, Guangxin Tao, Yang Liu, Chengjie Wang

122 Towards High-fidelity Artistic Image Vectorization via Texture-Encapsulated Shape Parameterization, Ye Chen, Bingbing Ni, Jinfan Liu, Xiaoyang Huang, Xuanhong Chen

123 Learning Vision from Models Rivals Learning Vision from Data, Yonglong Tian, Lijie Fan, KaiFeng Chen, Dina Katabi,
Dilip Krishnan, Phillip Isola 147
Efficient Multitask Dense Predictor via Binarization, Yuzhang Shang, Dan Xu, Gaowen Liu, Ramana Rao Kompella, Yan Yan 145
ReViT: Revisiting Mobile CNN From ViT Perspective, Ao Wang, Hui Chen, Ziia Lin, Jungong Han, Guiguang Ding 143
Enhancing Post-training Quantization Calibration through Contrastive Learning, Yuzhang Shang, Gaowen Liu, Ramana Rao Kompella, Yan Yan 140
FreeKD: Knowledge Distillation via Semantic Frequency Prompt, Yuan Zhang, Tao Huang, Jiaming Liu, Tao Jiang, Kuan Cheng, Shanghang Zhang 138
PTQ4SAM: Post-Training Quantization for Segment Anything, Chengtao Lv, Hong Chen, Jinyang Guo, Yifu Ding, Xianglong Liu 136
CLIP-KD: An Empirical Study of CLIP Model Distillation, Chuangyang Yang, Zhihui An, Libo Huang, Junyu Bi, Xingyi Yu, Han Yang, Boyu Diao, Yongjun Xu 134
Scaled Decoupled Distillation, Shicai Wei, Chunbo Luo, Yeaho Wang, Song Guo 130
Self-Distilled Masked Auto-Encoders are Efficient Video Anomaly Detectors, Nicolae-Cătălin Ristea, Florinel-Alin Crătoru, Radu Tudor Ionescu, Marius Popescu, Fahad Shahbaz Khan, Mubarak Shah 128
PikeLPN: Mitigating Overlooked Inefficiencies of Low-Precision Neural Networks, Marina Neseem, Conor McCollough, Randy Hsin, Chas Leichner, Shan Li, In Suk Chong, Andrew Howard, Lukasz Lew, Sherief Reda, Ville-Mikko Rautio, Daniele Moro 126
C2KD: Bridging the Modality Gap for Cross-Modal Knowledge Distillation, Fushuo Huan, Wenchao Xu, Jingcai Guo, Haozhao Wang, Song Guo 124
KD-DETR: Knowledge Distillation for Detection Transformer with Consistent Distillation Points Sampling, Yu Wang, Xin Li, Shengzhao Weng, Gang Zhang, Haixiao Yue, Haocheng Feng, Junyu Han, Enrui Ding 122
Towards Accurate Post-training Quantization for Diffusion Models, Changyuan Wang, Ziwei Wang, Xuwei Xu, Yansong Tang, Jie Zhou, Jiwen Lu 120
CURSOR: Scalable Mixed-Order Hypergraph Matching with CUR Decomposition, Qixuan Zheng, Ming Zhang, Hong Yan 118
Frozen Feature Augmentation for Few-Shot Image Classification, Andreas Bär, Neil Houlsby, Mostafa Dehghani, Manoj Kumar 116
Jointly Training and Pruning CNNs via Learnable Agent Guidance and Alignment, Alireza Ganjdanesh, Shangqian Gao, Heng Huang 114
Zero-TPrune: Zero-Shot Token Pruning through Leveraging of and Alignment, Jointly Training and Pruning CNNs via Learnable Agent Guidance and Alignment, Alireza Ganjdanesh, Shangqian Gao, Heng Huang 112
BilevelPruning: Unified Dynamic and Static Channel Pruning for Convolutional Neural Networks, Shangqian Gao, Yanfu Zhang, Feihu Huang, Heng Huang 110
Low-Rank Rescaled Vision Transformer Fine-Tuning: A Residual Design Approach, Wei Dong, Xing Zhang, Bihui Chen, Dawei Yan, Zhijun Lin, Qingsen Yan, Peng Wang, Yang Yang 108
EfficientSAM: Leveraged Masked Image Pretraining for Efficient Segment Anything, Yuyang Xiong, Bala Varadarajan, Lemeng Wu, Xiaoyu Xiang, Fanji Xiao, Chenchen Zhu, Xiao Liang Dai, Dilin Wang, Fei Sun, Forrestiland, Raghuraman Krishnamoorthi, Vikas Chandra 106
FlashEval: Towards Fast and Accurately Evaluation of Text-to-image Diffusion Generative Models, Lin Zhao, Tianchen Zhao, Zinlin Lin, Xuefei Ning, Guohao Dai, Huazhong Yang, Yu Wang 104
Instance-Aware Group Quantization for Vision Transformers, Xuefei Ning, Guohao Dai, Huazhong Yang, Yu Wang 102
Finding Lottery Tickets in Vision Models via Data-driven Spectral Foresight Pruning, Leonardo Iurado, Marco Ciccone, Tatiana Tommasi 100
Joint-Task Regularization for Partially Labeled Multi-Task Learning, Kento Nishi, Junsik Kim, Wanhua Li, Hanspeter Pfister 98
Auto-Train-Once: Controller Network Guided Automatic Network Pruning from Scratch, Xidong Wu, Shangqian Gao, Zeyu Zhang, Zhihui Lin, Runxue Bao, Yanfu Zhang, Xiaoqian Wang, Heng Huang 96
Reg-PTQ: Regression-specialized Post-training Quantization for Fully Quantized Object Detector, Yifu Ding, Weilun Feng, Chuyan Chen, Jinyang Guo, Xianglong Liu 94
MULTIFLOW: Shifting Towards Task-Agnostic Vision-Language Pruning, Matteo Farina, Massimiliano Mancini, Elia Cunegatti, Gaowen Liu, Giovanni Iacca, Elisa Ricci 92
MTLORA: Low-Rank Adaptation Approach for Efficient Multi-Task Learning, Ahmed Agiza, Marina Neseem, Sherief Reda 90
Resource-Efficient Transformer Pruning for Finetuning of Large Models, Fatih Ilhan, Gong Su, Selim Furkan Tekin, Tiansheng Huang, Shihao Hu, Ling Liu 88
Promptable Behaviors: Personalizing Multi-Objective Rewards from Human Preferences, Minyoung Hwang, Luca Weihs, Chanwoo Park, Kimin Lee, Aniruddha Kembhavi, Kianna Ehsani 86
Holodeck: Language Guided Generation of 3D Embodied AI Environments, 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 84
PhyScene: Physically Interactale 3D Scene Synthesis for Embodied AI, Yandong Yang, Xingjiang Jia, Peiyuan Zhi, Siyuan Huang, Zeng Kuan, Kunal Pratap Singh, Yejin Kim, Winson Han, Alvaro Herrasti, Ranjay Krishna, Dustin Schwenk, Eli VanderBilt, Aniruddha Kembhavi 82
OVER-NAV: Elevating Iterative Vision-and-Language Navigation with Open-Vocabulary Detection and StructureD Representation, Genlong Zhao, Guanbin Li, Weikai Chen, Yizhu Yu 78
MPS: A Multi-modal Open-ended Embodied System in Minecraft via Active Perception, Yiran Qin, Enshen Zhou, Qichang Liu, Zhenfei Yin, Lu Sheng, Ruimao Zhang, Yu Qiao, Jing Shao 76
Volumetric Environment Representation for Vision-Language Navigation, Rui Liu, Wenguang Wang, Yi Yang 74
Instance-aware Exploration-Verification-Exploitation for Instance ImageGoal Navigation, Xiaohao Lei, Min Wang, Wengang Zhou, Li Li, Houqiang Li 72
UniGarmentManip: A Unified Framework for Category-Level Garment Manipulation via Dense Visual Correspondence, Ruihai Wu, Haoran Lu, Yiyan Wang, Yubo Wang, Hao Dong 70
Evidential Active Recognition: Intelligent and Prudent Open-World Embodied Perception, Lei Fan, Mingfu Liang, Yunxuan Li, Gang Hua, Ying Wu 68
GenH2R: Learning Generalizable Human-to-Robot Handover via Scalable Simulation Demonstration and Imitation, Zifan Wang, Junyu Chen, Ziqing Chen, Pengwei Xie, Rui Chen, Li Yi 66
GOAT-Bench: A Benchmark for Multi-Modal Lifelong Navigation, Mukul Khanna, Ram Ramakrisha, Gunjan Chhablani, Sriman Venamamadra, Theophile Gervet, Matthew Chang, Zsolt Kira, Devendra Singh Chaplot, Dhruv Batra, Roobeh Mottaghi 64
Open Active-Vocabulary Recognition: Let Intelligent Moving Mitigate CLIP Limitations, Lei Fan, Jianxiong Zhou, Xiaoying Xing, Ying Wu 60
171 Rapid Motor Adaptation for Robotic Manipulator Arms, Yichao Liang, Kevin Ellis, João Henriques

172 Imagine Before Seeing: Self-Supervised Generative Map for Object Goal Navigation, Sixian Zhang, Xinyao Yu, Xinhong Song, Xiaohan Wang, Shuqiang Jang


174 GenNBV: Generalizable Next-Best-View Policy for Active 3D Reconstruction, Xiao Chen, Quanyi Li, Tai Wang, Tianfan Xue, Jiamingao Pang

175 An Interactive Navigation Method with Effect-oriented Affordance, Xiaohan Wang, Yuehu Liu, Xinhong Song, Yuji Liu, Sixian Zhang, Shuqiang Jang

176 A Category Agnostic Model for Visual Rearrangement, Yuyi Liu, Xinhong Song, Weijie Li, Xiaohan Wang, Shuqiang Jang

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

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


180 Model Adaptation for Time Constrained Embodied Control, Jaehyun Song, Minjung Yoo, Honguk Woo

181 EgoGen: An Egocentric Synthetic Data Generator, *Gen Li, Kai Feng Zhao, Siwei Zhang, Xiaohong Lyu, Mihai Dusmanu, Yan Zhang, Marc Pollefeys, Sinyu Tang

182 RoHM: Robust Human Motion Reconstruction via Diffusion, Siwei Zhang, Bharat Lal Bhagatkar, Yuanlu Xu, Alexander Winkler, Petr Kadlec, Sinyu Tang, Federica Bogo

183 An N-Point Linear Solver for Line and Motion Estimation with Event Cameras, Ling Gao, Daniel Gehrig, Hang Su, Davide Scaramuzza, Laurent Kneip

184 A Subspace-Constrained Yttrium’s Estimator and its Applications to Structure from Motion, *Feng Yu, Teng Zhang, Gilad Lerman

185 SpiderMatch: 3D Shape Matching with Global Optimality and Geometric Consistency, Paul Roetzer, Florian Bernard

186 You’ll Never Walk Alone: A Sketch and Text Duet for Fine-grained Object Retrieval, Fanjie Kong, Yanbei Chen, Jianrui Cai, Davide Modolo

187 CricaVPR: Cross-image Correlation-aware Representation Learning for Visual Place Recognition, Feng Lu, Xiangyan Lan, Lijun Zhang, Dongmei Jiang, Yaowei Wang, Chen Zhang, Yuting Zhang, Siliang Tang

188 Hyperbolistic Learning with Synthetic Captions for Open-World Detection, Fanjie Kong, Yanbei Chen, Jianrui Cai, Davide Modolo

189 ProTeCt: Prompt Tuning for Taxonomic Open Set Classification, Tz-Ying Wu, Chih-Hui Ho, Nuno Vasconcelos

190 CAT: Exploiting Inter-Class Dynamics for Domain Adaptive Object Detection, Mikhail Kennerley, Jian-Gang Wang, Bharadwaj Veeravalli, Robby T. Tan

191 Text Is MASS: Modeling as Stochastic Embedding for Text-Video Retrieval, Jiamian Wang, Guohao Sun, Pichao Wang, Dongfang Liu, Sohail Dianat, Majid Rabbani, Raghuram Rao, Zhenglu Tao

192 UniMODE: Unified Monocular 3D Object Detection, Zhi-Fan Wu, Zhanlong Zhuo, Zhiqiang Shen

193 OVMR: Open-Vocabulary Recognition with Multi-Modal Features, Zehong Ma, Shilhang Zhang, Longhui Wei, Qi Tian

194 From Isolated Islands to Pangea: Unifying Semantic Space for Human Action Understanding, Yong-Lu Li, Xiaoqian Wu, Xinpeng Liu, Zehao Wang, Yiming Dou, Yunji Kun, Junyi Zhang, Xiyong Li, Xudong Lu, Jingru Tan, Ce Wu Lu

195 Language-conditioned Detection Transformer, Jang Hyun Cho, Philipp Krähenbühl

196 Distribution-aware Knowledge Prototyping for Non-exemplar Lifelong Person Re-identification, Kunlun Xu, Xu Zou, Yuxin Peng, Jiahuan Zhou

197 Learning Continual Compatible Representation for Re-indexing Free Lifelong Person Re-identification, Zhenyu Cui, Jiahuan Zhou, Xun Wang, Manyu Zhu, Yuxin Peng

198 Active Object Detection with Knowledge Aggregation and Distillation from Large Models, Dejie Yang, Yang Liu

199 SHiNe: Semantic Hierarchy Nexus for Open-vocabulary Object Detection, Mingxuan Liu, Tyler L. Hayes, Elisa Ricci, Gabriela Csurka, Riccardo Volpi

200 Object Recognition as Next Token Prediction, Kaiyu Yue, Bor-Chun Chen, Jonas Geiping, Hengduo Li, Tom Goldstein, Ser-Nam Lim

201 Exploring the Potential of Large Foundation Models for Open-Vocabulary HOI Detection, Ting Lei, Shaofeng Yin, Yang Liu

202 Gradient Reweighting: Towards Imbalanced Class-Incremental Learning, Jiageng He

203 Multi-View Attentive Contextualization for Multi-View 3D Object Detection, Xianpeng Liu, Ce Zheng, Ming Qian, Nan Xue, Chen Chen, Zhebin Zhang, Chen Li, Tianfu Wu

204 RealNet: A Feature Selection Network with Realistic Synthetic Anomaly for Anomaly Detection, Ximiao Zhang, Min Xu, Xuzhuang Zhou

205 Generalized Large-Scale Data Condensation via Various Backbone and Statistical Matching, Shitong Shao, Zeyuan Yin, Muxin Zhou, Xiongding Zhang, Zhiquan Shen

206 Unleashing Unlabeled Data: A Paradigm for Cross-View Geo-Localization, Guopeng Li, Ming Qian, Gui-Song Xia

207 PointOBB: Learning Oriented Object Detection via Single Point Supervision, Junwei Luo, Xue Yang, Yi Yu, Qingyun Li, Junchi Yan, Yansheng Li

208 Scene-adaptive and Region-aware Multi-modal Prompt for Open Vocabulary Object Detection, Xiaowei Zhao, Xianglong Liu, Duorui Wang, Yajun Gao, Zhide Liu

209 Revisiting the Domain Shift and Sample Uncertainty in Multi-source Active Domain Transfer, Wenqiao Zhang, Zheqi Lv, Hao Zhou, Jia-Wei Liu, Juncheng Li, Mengze Li, Yunfei Li, Dongping Zhang, Yueting Zhang, Siliang Tang

210 Hyperbolistic Learning with Synthetic Captions for Open-World Detection, Fanjie Kong, Yanbei Chen, Jianrui Cai, Davide Modolo

211 CricaVPR: Cross-image Correlation-aware Representation Learning for Visual Place Recognition, Feng Lu, Xiangyan Lan, Lijun Zhang, Dongmei Jiang, Yaowei Wang, Chun Yuan

212 Point2RBox: Combine Knowledge from Synthetic Visual Patterns for End-to-end Oriented Object Detection with Single Point Supervision, Yi Yu, Xue Yang, Qingyun Li, Feiping Da, Lifeng Dai, Yu Qiao, Junchi Yan

213 Scene Adaptive Sparse Transformer for Event-based Object Detection, Yansong Peng, Hebei Li, Yueyi Zhang, Xiaoyan Sun, Feng Wu

214 Visual Delta Generator with Large Multi-modal Models for Semi-supervised Composed Image Retrieval, Young Kyun Jang, Donghyun Kim, Zhanhong Meng, Dat Huynh, Ser-Nam Lim

215 Preserving Fairness Generalization in Deepfake Detection, Li Lin, Xihan He, Yan Ju, Xin Wang, Feng Ding, Shu Hu

216 Text-to-Image Diffusion Models are Great Sketch-Photo Matchmakers, Subhadeep Koley, Ayan Kumar Bhunia, Aneeshaan Sain, Pinaki Nath Chowdhury, Tao Xiang, Yi-Zhe Song

217 PromptAD: Learning Prompts with only Normal Samples for Few-Shot Anomaly Detection, Xiaofan Li, Zhihong Zhang, Xin Tan, Chengwei Chen, Yanyun Qu, Yuan Xie, Lihuang Ma

218 Structured Model Probing: Empowering Efficient Transfer Learning by Structured Regularization, Zhi-Fan Wu, Chaojie Mao, Wue Wang, Jianwen Jiang, Yiliang Lv, Rong Jin

219 How to Handle Sketch-Abstraction in Sketch-Based Image Retrieval?, Subhadeep Koley, Ayan Kumar Bhunia, Aneeshaan Sain, Pinaki Nath Chowdhury, Tao Xiang, Yi-Zhe Song
Shallow-Deep Collaborative Learning for Unsupervised Visible-Infrared Person Re-Identification, Bin Yang, Jun Chen, Mang Ye
221 Solving the Catastrophic Forgetting Problem in Generalized Category Discovery, Xinzi Cao, Xiawu Zheng, Guanhong Wang, Weijiang Yu, Yunhang Shen, Ke Li, Yutong Lu, Yonghong Tian
222 Active Generalized Category Discovery, Shijie Ma, Fei Zhu, Zhun Zhong, Xu-Yao Zhang, Cheng-Lin Liu
223 YOLO-World: Real-Time Open-Vocabulary Object Detection, Tianheng Cheng, Lin Song, Yixiao Ge, Wenyu Li, Xinggang Wang, Ying Shan
224 Theoretically Achieving Continuous Representation of Oriented Bounding Boxes, Zikai Xiao, Guoyu Yang, Xue Yang, Taijiang Mu, Junchi Yan, Shimin Hu
225 Decoupled Pseudo-labeling for Semi-Supervised Monocular 3D Object Detection, Jiacheng Zhang, Jiaming Li, Xiangru Liu, Wei Zhang, Xiao Tan, Junyu Han, Enru Ding, Jingdong Wang, Guanlin Li
226 LEOD: Label-Efficient Object Detection for Event Cameras, Ziyi Wu, Mathias Gehrig, Qing Lyu, Xudong Liu, Igor Gillitschenski
227 Lane2Seq: Towards Unified Lane Detection via Sequence Generation, Kungyang Zhou
228 Open-World Human-Object Interaction Detection via Multi-modal Prompts, Jie Yang, Bingliang Li, Ailing Zeng, Lei Zhang, Ruimao Zhang
229 DETRs Beat YOLOS on Real-time Object Detection, Yian Zhao, Wenyu Lv, Shangliang Xu, Jinnan Wei, Guanzhong Wang, Qingqiang Dang, Yi Liu, Jie Chen
230 Exploring Region-Word Alignment in Built-in Detector for Open-Vocabulary Object Detection, Heng Zhang, Qiuyi Zhao, Linyu Zheng, Hao Zhei, Zhiwei Ge, Tianhao Li, Sulong Xu
231 Referring Expression Counting, ✻ Siyuan Dai, Jun Liu, Ngai-Man Cheung
232 ActiVE-DC: Distribution Calibration for Active Finetuning, Wenhua Xu, Zhenghui Hu, Yu Lu, Jinzhou Meng, Qingjie Liu, Yunhong Wang
233 LaRE²: Latent Reconstruction Error Based Method for Diffusion-Generated Image Detection, Yunpeng Luo, Junlong Du, Ke Yan, Shouhong Dang
234 Fine-grained Prototypical Voting with Heterogeneous Mixup for Semi-supervised 2D-3D Cross-modal Retrieval, Fan Zhang, Xian-Sheng Hua, Chong Chen, Xiao Luo
235 MS-DETR: Efficient DETR Training with Mixed Supervision, Chuyang Zhao, Yifan Sun, Wenhao Wang, Qiang Chen, Enru Ding, Yi Yang, Jingdong Wang
236 Context-based and Diversity-driven Specificity in Compositional Zero-Shot Learning, Yun Li, Zhe Liu, Hang Chen, Lina Yao
237 Pixel-level Semantic Correspondence through Layout-aware Representation Learning and Multi-scale Matching Integration, Yixuan Sun, Zhangyue Yin, Haibo Wang, Yan Wang, Xipeng Qiu, Weifeng Ge, Wengiang Zhang
238 Exploiting Inter-sample and Inter-feature Relations in Dataset Distillation, Wenxia Deng, Wenbin Li, Tianyu Ding, Lei Wang, Hongqiang Zhang, Kuihua Huang, Jing Hao, Yang Gao
239 Point Segment and Count: A Generalized Framework for Object Counting, Zhizhong Huang, Mingliang Dai, Yi Zhang, Junping Zhang, Hongming Shan
240 Dual Pose-invariant Embeddings: Learning Category and Object-specific Discriminative Representations for Recognition and Retrieval, Rohan Sarkar, Avinash Kak
241 Riemannian Multinomial Logistics Regression for SPD Neural Networks, Zheng Chen, Yue Song, Gaowen Liu, Ramana Rao Kompella, Xiao-Jun Wu, Nicu Sebe
242 Learning for Transductive Threshold Calibration in Open-World Recognition, Qin Zhang, Dongsheng An, Tianjun Xiao, Tong He, Qingming Tang, Ying Nian Wu, Joseph Tighe, Yifan Xing
243 Region-Based Representations Revisited, Michal Shlapentokh-Rothman, Ansel Blume, Yao Xiao, Yuxun Wu, Sutharanan TV, Heyi Tao, Jie Yang Lee, Wilfredo Torres, Yu-Xiong Wang, Derek Hoiem
244 Magic Tokens: Select Diverse Tokens for Multi-modal Object Re-Identification, Pingzhang Zheng, Yuhao Wang, Yang Liu, Zhengzheng Tu, Huchuan Lu
245 Harnessing the Power of MLLMs for Transferable Text-to-Image Person ReID, Wentao Tan, Changxing Ding, Jiayu Jiang, Fei Wang, Yibing Zhan, Dapeng Tao
246 Holistic Features are almost Sufficient for Text-to-Video Retrieval, Kaibin Tian, Ruixiang Zhao, Zijie Xin, Bangxiang Lan, Xirong Li
247 Enhancing the Power of OOD Detection via Sample-Aware Model Selection, Feng Xue, Zi He, Yuan Zhang, Chuanlong Xie, Zhenqiu Li, Falong Tan
248 PanoOCC: Unified Occupancy Representation for Camera-based 3D Panoptic Segmentation, Yuni Wang, Yuntao Chen, Xingyu Liao, Lue Fan, Zhaoxiang Zhang
249 VSCode: General Visual Salient and Camouflaged Object Detection with 2D Prompt Learning, Ziyang Luo, Nian Liu, Wangbo Zhao, Xuguang Yang, Dingwen Zhang, Deng-Ping Fan, Fahad Khan, Junwei Han
250 D3still: Decoupled Differential Distillation for Asymmetric Image Retrieval, Yi Xie, Yihong Lin, Wenjie Cai, Xuemiao Xu, Huaidong Zhang, Yong Du, Shengfeng He
251 SPOD: Spiking Fusion Object Detector, Yimeng Fan, Wei Zhang, Changsong Liu, Mingyang Li, Wenrui Lu
252 Depth-Aware Conceived Crop Detection in Dense Agricultural Scenes, Liqiong Wang, Jinyu Yang, Yanlu Zhang, Fangyi Wang, Feng Zheng
253 Extreme Point Supervised Instance Segmentation, Hyeonjin Lee, Sehyun Hwang, Suha Kwak
254 Enhance Image Classification via Inter-Class Image Mixup with Diffusion Model, Zhicai Wang, Longhui Wei, Tan Wang, Heyu Chen, Yanbin Hao, Xiang Wang, Xiangnan He, Qi Tian
255 Multimodal Industrial Anomaly Detection by Crossmodal Feature Mapping, Alex Costanzino, Pierluigi Zama Ramirez, Giuseppe Lisanti, Luigi Di Stefano
256 Effective Video Mirror Detection with Inconsistent Motion Cues, Alex Warren, Ke Xu, Jaying Lin, Gary K.L. Tam, Ryan W.H. Lau
257 Multi-Attribute Interactions Matter for 3D Visual Grounding, Can Xu, Yuehui Han, Rui Xu, Le Hui, Jin Xie, Jian Yang
258 Looking 3D: Anomaly Detection with 2D-3D Alignment, Ankan Bhunia, Chandian Li, Hakan Bilen
259 Characteristics Matching Based Hash Codes Generation for Efficient Fine-grained Image Retrieval, Zhen-Duo Chen, Li-Jun Zhao, Zi-Chao Zhang, Xin Luo, Xin-Shun Xu
260 EASE-DETR: Easing the Competition among Object Queries, Yulu Gao, Yifan Sun, Xudong Ding, Chuyang Zhao, Si Liu
261 ProS: Prompting-to-simulate Generalized knowledge for Universal Cross-Domain Retrieval, Kaipeng Fang, Jingkuan Song, Lianli Gao, Pengpeng Zeng, Zhi-Qi Cheng, Xiyao Li, Heng Tao Shen
262 Exploring Orthogonality in Open World Object Detection, Zhicheng Sun, Jinghan Li, Yadong Mu
263 A Generative Approach for Wikipedia-Scale Visual Entity Recognition, Mathilde Caron, Ahmet Iscen, Alireza Fathi, Cordelia Schmid
264 Unleashing Channel Potential: Space-Frequency Selection Convolutions for SAR Object Detection, Ke Li, Di Wang, Zhangyuan Hu, Wenxuan Zhu, Shaofeng Li, Quan Wang
265 Hyperspherical Classification with Dynamic Label-to-Prototype Assignment, Mohammad Saeed Ebrahimii Saadabadi, Ali Babouei, Sahar Rahimi Malakshar, Nasser M. Nasrabadi
266 A Pedestrian is Worth One Prompt: Towards Language Guidance ✻ Person Re-Identification, Zexian Yang, Dayan Wu, Chenming Wu, Zheng Lin, Jingzi Gu, Weiping Wang
267 VSRD: Instance-Aware Volumetric Silhouette Rendering for Weakly Supervised 3D Object Detection, Zhua Liu, Hiroki Sakuma, Masatoshi Okutomi
268 Improving Visual Recognition with Hyperbolic Visual Hierarchy Mapping, Hyeongjun Kwon, Jinhyun Jang, Jin Kim, Kwonyoung Kim, Kwanghoon Sohn
269 On Train-Test Class Overlap and Detection for Image Retrieval, Chull Hwer Song, Jooyoung Yoon, Taebaek Hwang, Shunghyun Choi, Yeong Hyeon Gu, Yannis Avrithis
270 Multi-Scale Video Anomaly Detection by Multi-Grained ✻ Spatio-Temporal Representation Learning, Menghao Zhang, Jingyu Wang, Qi Qi, Hai Feng Sun, Zirui Zhuang, Pengfei Ren, RuiLong Ma, Jianxin Liao

272 Rethinking Boundary Discontinuity Problem for Oriented Object Detection, Hang Xu, Xinyuan Liu, Haoran Xu, Yike Ma, Zunjie Zhu, Chenggang Yan, Feng Dai

273 Hybrid Proposal Refiner: Revisiting DETR Series from the Faster R-CNN Perspective, Jinjing Zhao, Fangyuan Wei, Chang Xu

274 Retrieval-Augmented Open-Vocabulary Object Detection, Jooyeon Kim, Eurlang Cho, Sehyung Kim, Hyunwoo J. Kim

275 LIAR-based Person Re-identification, Wexuan Guo, Zhiyu Pan, Yingping Liang, Ziheng Xi, Zhicheng Zhang, Jianjiang Feng, Jie Zhou

276 EventDance: Unsupervised Source-free Cross-modal Adaptation for Event-based Object Recognition, Xu Zheng, Lin Wang

277 All in One Framework for Multimodal Re-identification in the Wild, He Li, Mang Ye, Ming Zhang, Bo Du

278 Logarithmic Lenses: Exploring Log RGB Data for Image Classification, Bruce A. Maxwell, Sumegha Singhania, Avmish Patel, Rahul Kumar, Heather Frying, Sihan Li, Haoran Sun, Ping He, Zewen Li

279 ID-like Prompt Learning for Few-Shot Out-of-Distribution Detection, Yichen Bai, Zongbo Han, Bing Cao, Xiaoheng Jiang, Qinghua Hu, Changqing Zhang

280 Infrared Small Target Detection with Scale and Location Sensitivity, Qiankun Liu, Rui Liu, Bolun Zheng, Hongkui Wang, Ying Fu

281 SURE: Survey Recipes for building reliable and robust deep networks, Yuting Li, Yingyi Chen, Xuanlong Yu, Dexiong Chen, Xi Shen

282 Hyperbolic Anomaly Detection, Huimin Li, Zhentao Chen, Yunhao Xu, Junlin Hu

283 Instruct-RelD: A Multi-purpose Person Re-identification Task with Instructions, Weizhen He, YiHeng Deng, Shixiang Tang, Qihao Chen, Qingsong Xie, Yizhou Wang, Lei Bai, Feng Zhu, Rui Zhao, Wanli Ouyang, Donglian Qi, Yunfeng Yan

284 CA-Jaccard: Camera-aware Jaccard Distance for Person Re-identification, Yiyun Chen, Zheyi Fan, Zhaorui Chen, Xiyuan Zhu

285 Improved Zero-Shot Classification by Adapting VLMs with Text Descriptions, Oindrila Saha, Grant Van Horn, Subhransui Majumdar


287 Supervision, Atom-Level Optical Chemical Structure Recognition with Limited Optimal Transport Aggregation for Visual Place Recognition, Ziteng Gao, Zhan Tong, Kevin Qinghong Lin, Joya Chen, Mike Zheng Shou

288 Not All Classes Stand on Same Embeddings: Calibrating a Semantic Distance with Metric Tensor, Jae Hyeon Park, Gyoomin Lee, Seunggi Park, Sung In Cho

289 Improving Single Domain-Generalized Object Detection: A Focus on Diversification and Alignment, Muhammad Sohail Danish, Muhammad Haris Khan, Muhammad Akhtar Munir, M. Saquib Sarfraz, Mohsen Ali

290 On the Estimation of Image-matching Uncertainty in Visual Place Recognition, Mubarak Zaffar, Liangliang Nan, Julian F. P. Kooij

291 Supervised Anomaly Detection for Complex Industrial Images, Aimira Baieber, David Hurych, Victor Besnier, Olivier Bernard

292 Towards Generalist Anomaly Detection via In-context Residual Learning with Few-shot Sample Prompts, Jiawen Zhu, Choubo Ding, Yu Tian, Guansong Pang

293 Learning to Navigate Efficiently and Precisely in Real Environments, Guillaume Bono, Hervé Poirier, Leonid Antsfeld, Gianluca Monaci, Boris Chidlovskii, Christian Wolf

294 Task-Conditioned Adaptation of Visual Features in Multi-Task Policy Learning, Pierre Marza, Laetitia Matignon, Olivier Simondon, Christian Wolf

295 FastMAC: Stochastic Spectral Sampling of Correspondence Graph, Yifei Zhang, Hao Zhao, Hongyang Li, Sheng Chen

296 FoundationPose: Unified 6D Pose Estimation and Tracking of Novel Objects, Bowen Wen, Wei Yang, Jan Kautz, Stan Birchfield

297 SAGE: Controllable Articulation GEneration, Jiayi Liu, Hou In Yam Tam, Ali Mahdavi-Amiri, Manolis Savva

298 SingularTrajectory: Universal Trajectory Predictor Using Diffusion Model, Inhwon Bae, Young-Jae Park, Hae-Gon Jeon

299 Language-driven Grasp Detection, An Dinh Vuong, Minh Nhat Vu, Baoru Huang, Nghia Nguyen, Hieu Le, Thieu Vo, Anh Nguyen

300 MemoNav: Working Memory Model for Visual Navigation, Hongxin Li, Zeyu Wang, Xu Yang, Yuran Yang, Shuqi Mei, Zhaoxiang Zhang

301 NOPE: Novel Object Pose Estimation from a Single Image, Van Nguyen Nguyen, Thibault Groueix, Georgy Pominatkin, Yinlin Hu, Renaud Marlet, Mathieu Salzmann, Vincent Lepetit

302 Dexterous Grasp Transformer, Guo-Hao Xu, Yi-Lin Wei, Dian Zheng, Xiao-Ming Wu, Wei-Shi Zheng

303 Versatile Navigation Under Partial Observability via Value-guided Diffusion Policy, Gengyu Zhang, Hao Tang, Yan Yan

304 CyberDemo: Augmented Simulated Human Demonstration for Real-World Dexterous Manipulation, Jun Wang, Yuzhe Qin, Kaiming Kuang, Yigit Korkmaz, Akhilan Gurumoorthy, Hao Su, Xiaolong Wang

Yves Moreau

298 Novel Class Discovery for Ultra-Fine-Grained Visual Categorization, Yu Liu, Yaqi Cai, Qi Jia, Binglin Qiu, Weimian Wang, Yan Pu

299 Attribute-Guided Pedestrian Retrieval: Bridging Person Re-ID with Internal Attribute Variability, Yan Huang, Zhang Zhang, Qiang Wu, Yi Zhong, Liang Wang

300 Robust Noisy Correspondence Learning with Equivariant Similarity Consistency, Yucheng Yang, Likai Wang, Erkun Yang, Cheng Deng

301 Bootstrapping Similarity-Formers from Vision Foundation Models, Ziteng Gao, Zhan Tang, Kevin Qinghong Lin, Joya Chen, Mike Zheng Shou

302 Improving Single Domain-Generalized Object Detection: A Focus on Diversification and Alignment, Muhammad Sohail Danish, Muhammad Haris Khan, Muhammad Akhtar Munir, M. Saquib Sarfraz, Mohsen Ali

303 On the Estimation of Image-matching Uncertainty in Visual Place Recognition, Mubarak Zaffar, Liangliang Nan, Julian F. P. Kooij

304 Supervised Anomaly Detection for Complex Industrial Images, Aimira Baieber, David Hurych, Victor Besnier, Olivier Bernard

305 Fourier-basis Functions to Bridge Augmentation Gap: Rethinking Frequency Augmentation in Image Classification, Puru Vaish, Shunxin Wang, Nicola Strisciuglio

306 TransNeXt: Robust Foveal Visual Perception for Vision Transformers, Dai Shi

307 Plug and Play Active Learning for Object Detection, Chenhongyi Yang, Lichao Huang, Elliot J. Crowley

308 BoQ: A Place is Worth a Bag of Learnable Queries, Amir Ali-bey, Brahim Chaib-draa, Philippe Giguère

309 From Coarse to Fine-Grained Open-Set Recognition, Nico Lang, Vésteinn Snæbjarnarson, Elejah Cole, Oisin Mac Aodha, Christian Igel, Serge Belongie

310 Exploring Pose-Aware Human-Object Interaction via Hybrid Learning, Eastman Z Y Wu, Yali Li, Yuan Wang. Shengjinnang Wang

311 Toward Generalist Anomaly Detection via In-context Residual Learning with Few-shot Sample Prompts, Jiawen Zhu, Guansong Pang

312 Learning to Navigate Efficiently and Precisely in Real Environments, Guillaume Bono, Hervé Poirier, Leonid Antsfeld, Gianluca Monaci, Boris Chidlovskii, Christian Wolf

313 Task-Conditioned Adaptation of Visual Features in Multi-Task Policy Learning, Pierre Marza, Laetitia Matignon, Olivier Simondon, Christian Wolf

314 FastMAC: Stochastic Spectral Sampling of Correspondence Graph, Yifei Zhang, Hao Zhao, Hongyang Li, Sheng Chen

315 FoundationPose: Unified 6D Pose Estimation and Tracking of Novel Objects, Bowen Wen, Wei Yang, Jan Kautz, Stan Birchfield
PROGRAM GUIDE

THURSDAY, JUNE 20

325 SchurVINS: Schur Complement-Based Lightweight Visual Inertial Navigation System, Yunfei Fan, Tianyu Zhao, Guidong Wang

326 READ: Retrieval-Enhanced Asymmetric Diffusion for Motion Planning, Takeku Oba, Matthew Walter, Norimichi Ukita

327 Retrieval-Augmented Embodied Agents, Yichen Zhu, Zhical Ou, Xiaofeng Mou, Jian Tang

328 Collaborative Semantic Occupancy Prediction with Hybrid Feature Fusion in Connected Automated Vehicles, Rui Song, Chenwei Liang, Huo Cao, Zhiran Yan, Walter Zimmer, Markus Gross, Andreas Festag, Alois Knoll

329 Diffusion-EDFs: Bi-equivariant Denoising Generative Modeling on SE(3) for Visual Robotic Manipulation, Hyunwoo Ryu, Jiwoo Kim, Hyunseok An, Junwoo Chang, Joohwan Seo, Taehan Kim, Yubin Kim, Chaewon Hwang, Jongeun Choi, Roberto Horowitz

330 Adaptive VIO: Deep Visual-inertial Odometry with Online Continual Learning, Yoyu Pan, Wugen Zhou, Yingdian Cao, Hongbin Zha

331 F3Loc: Fusion and Filtering for Floorplan Localization, Changan Chen, Rui Wang, Christoph Vogel, Marc Pollefeys

332 Gaussian Splitting SLAM, Hidenobu Matsuki, Riku Murai, Paul H.J. Kelly, Andrew J. Davison

333 SUGAR: Pre-training 3D Visual Representations for Robotics, Shizhe Chen, Ricardo Garcia, Ivan Laptev, Cordelia Schmid

334 ManiLLM: Embodied Multimodal Large Language Model for Object-Centric Robotic Manipulation, Xiaoli Li, Mingxu Zhang, Yiran Geng, Haoaren Geng, Yuxing Long, Yan Shen, Renrui Zhang, Jiaoming Liu, Hao Dong

335 Open-Vocabulary Object 6D Pose Estimation, Jaime Corsetti, Davide Boscaiini, Changjue Oh, Andrea Cavallaro, Fabio Poiesi

336 Hierarchical Diffusion Policy for Kinematics-Aware Multi-Task Robotic Manipulation, Xiaoxa Ma, Sumit Patidar, Iain Haughton, Stephen James

337 Smart Help: Strategic Opponent Modeling for Proactive and Adaptive Robot Assistance in Households, Zhichao Cao, Zidong Wang, Siwen Xie, Anji Liu, Lifeng Fan

338 Generalizing 6-DoF Grasp Detection via Domain Prior Knowledge, Haoxiang Ma, Modi Shi, Boyang Gao, Di Huang

339 A Simple and Effective Point-based Network for Event Camera 6-DOFs Pose Relocalization, Hongwei Ren, Jadong Zhu, Yue Zhou, Haotian Fu, Yulong Huang, Bojun Cheng

340 Neural Visibility Field for Uncertainty-Driven Active Mapping, Shangjie Xue, Jesse Dill, Pranay Mathur, Frank Dellaert, Panagiotis Tsiotra, Danfei Xu

341 SPIN: Simultaneous Perception Interaction and Navigation, Shagun Uppal, Anany Agarwal, Haoyu Xiong, Kenneth Shaw, Deepak Pathak

342 SceneFun3D: Fine-Gained Functionality and Affordance Understanding in 3D Scenes, Alexandros Delitzas, Ayca Takmaz, Federico Tombari, Robert Sumner, Marc Pollefeys, Francis Engelmann

343 PredToken: Predicting Unknown Tokens and Beyond with Coarse-to-Fine Iterative Decoding, Xuesong Nie, Haoyuan Jin, Yunfeng Yan, Xi Chen, Zhizhang Zhu, Donglian Qi

344 TIM: A Time Interval Machine for Audio-Visual Action Recognition, Jacob Chalk, Jaesung Huh, Evangelos Kazakos, Andrew Zisserman, Dima Damen

345 AutoAD III: The Prequel – Back to the Pixels, Tengda Han, Max Bain, Arsha Nagrani, Gul Varol, Weidi Xie, Andrew Zisserman


347 Progress-Aware Online Action Segmentation for Egocentric Procedural Task Videos, Yuhai Shen, Ehsan Elhamifar

348 Video ReCap: Recursive Captioning of Hour-Long Videos, Md Mohaiminul Islam, Ngaon Ha, Xitong Yang, Tushar Nagarajan, Lorenzo Torresani, Gedas Bertasius


350 MovieChat: From Dense Token to Sparse Memory for Long Video Understanding, Enxin Song, Wenhao Chai, Guanhong Wang, Yucheng Zhang, Haoyang Zhou, Feiyang Wu, Haoze Chi, Xin Guo, Tian Ye, Yanting Zhang, Yan Lu, Jeng-Neng Hwang, Gaogang Wang

351 Learning Group Activity Features Through Person Attribute Prediction, Chihiro Nakatani, Hiroaki Kawashima, Norimichi Ukita

352 Streaming Dense Video Captioning, Xingyi Zhou, Anurag Arhab, Shyamal Buch, Shen Yan, Austin Myers, Xuelian Xiong, Arsha Nagrani, Cordelia Schmid

353 Efficient and Effective Weakly-Supervised Action Segmentation via Action-Transition-Aware Boundary Alignment, Angchi Xu, Wei-Shi Zheng

354 Benchmarking the Robustness of Temporal Action Detection Models Against Temporal Corruptions, Runhao Zeng, Xiaoyong Chen, Jiaming Liang, Huisi Wu, Guangzhong Cao, Yong Guo

355 A Backpack Full of Skills: Egocentric Video Understanding with Diverse Task Perspectives, Simone Alberto Peirone, Francesca Pintilli, Antonio Allegro, Giuseppe Avera

356 Summarize the Past to Predict the Future: Natural Language Descriptions of Context Boost Multimodal Object Interaction Anticipation, Razvan-George Pasca, Alexey Gavryushin, Muhammad Hamza, Yen-Ling Kuo, Kaichun Mo, Luc Van Gool, Otmar Hilliges, Xi Wang

357 Open-Vocabulary Video Anomaly Detection, Peng Wu, Xuerong Zhou, Guansong Pang, Yujia Sun, Jing Liu, Peng Wang, Yanning Zhang

358 Task-Driven Exploration: Decoupling and Inter-Task Feedback for Joint Moment Retrieval and Highlight Detection, Jin Yang, Ping Wei, Huan Li, Ziyang Ren

359 Prompt-Enhanced Multiple Instance Learning for Weakly Supervised Video Anomaly Detection, Junxi Chen, Liang Li, Li Su, Zheng-jun Zha, Qingming Huang

360 Context-Guided Spatio-Temporal Video Grounding, Xin Gu, Heng Fan, Yan Huang, Tiejian Luo, Libo Zhang

361 Just Add ?! Pose Induced Video Transformers for Understanding Activities of Daily Living, Dominick Reilly, Srijan Das

362 Action Detection via an Image Diffusion Process, Lin Geng Foo, Tianqiao Li, Hossein Rahmani, Jun Liu

363 LLMs are Good Sign Language Translators, Jia Geng, Lin Geng Foo, Yixuan He, Hossein Rahmani, Jun Liu

364 End-to-End Spatio-Temporal Action Localisation with Video Transformers, Alexey A. Gritsenko, Xuehan Xiong, Josip Djolonga, Mostafa Dehghani, Chen Sun, Mario Lucic, Cordelia Schmid, Anurag Arhab

365 HIG: Hierarchical Interleaving Graph Approach to Scene Graph Generation in Video Understanding, Trong-Thuan Nguyen, Pha Nguyen, Khoa Luu

366 LLMs are Good Action Recognizers, Haoxuan Qu, Yujin Cai, Jun Liu

367 VideoLLM-online: Online Video Large Language Model for Streaming Video, Jaya Chen, Zhaoyang Lv, Shiwei Wu, Kevin Qinghong Lin, Chenan Song, Difei Gao, Jia-Wei Liu, Ziteng Gao, Dongxing Mao, Mike Zheng Shou

368 What When and Where? Self-Supervised Spatio-Temporal Grounding in Untrimmed Multi-Action Videos from Narrated Instructions, Brian Chen, Nina Shevetsova, Andrew Rouditchenko, Daniel Kendermann, Samuel Thomas, Shih-Fu Chang, Rogerio Feris, James Glass, Hilde Kuehne

369 Narrative Action Evaluation with Prompt-Guided Multimodal Interaction, Shiyi Zhang, Sule Bai, Guanyi Chen, Lei Chen, Jiwen Lu, Junle Wang, Yansong Tang

370 Realigning Confidence with Temporal Saliency Information for Point-Level Weakly-Supervised Temporal Action Localization, Ziyang Xia, Jian Cheng, Siyu Liu, Yongxiang Hu, Shiguang Wang, Yijie Zhang, Liwan Dang

371 Action-slot: Visual Action-centric Representations for Multi-label Atomic Activity Recognition in Traffic Scenes, Chi-Hsi Kung, Shu-Wei Lu, Yi-Huan Tsai, Yi-Ting Chen

372 LoCoNet: Long-Short Context Network for Active Speaker Detection, Xizi Wang, Feng Cheng, Gedas Bertasius

373 Neighbor Relations Matter in Video Scene Detection, Jiawei Tan, Hongxing Wang, Jiaxin Li, Zhilong Ou, Zhanqiang Qian

374 PREGO: Online Mistake Detection in PProcedural EGOCentric Videos, Alessandro Flaborea, Guido Maria D’Amelio di Melendugno, Leonardo Plini, Luca Scofano, Edoardo De Matteis, Antonio Furnari, Giovanni Maria Farinella, Fabio Galasso
Learning Object State Changes in Videos: An Open-World Perspective, Zhihu Xue, Kumar Ashutosh, Kristen Grauman
Enhanced Motion-Text Alignment for Image-to-Video Transfer Learning, Wei Zhang, Chaowu Wang, Tongliang Liu, Xinmei Tian, Xu Shen, Jieping Ye
Asymmetric Masked Distillation for Pre-Training Small Foundation Models, Zhiyu Zhao, Bingkun Huang, Sen Xing, Gangshan Wu, Yu Qiao, Limin Wang
Harnessing Large Models for Training-free Video Anomaly Detection, Luca Zanella, Willi Menapace, Massimiliano Mancini, Yiming Wang, Elisa Ricci
SportsHII: A Dataset for Human-Human Interaction Detection in Sports Videos, Tao Wu, Runyu He, Gangshan Wu, Limin Wang
ViT-C: Video-conditioned Text Representations for Activity Recognition, Kumara Khatapotipitya, Anurag Arnab, Arsha Nagrani, Michael S. Ryoo
Dual DETRs for Multi-Label Temporal Action Detection, Yuhua Zhu, Guozhen Zhang, Jing Tan, Gangshan Wu, Limin Wang
Adapting Short-Term Transformers for Action Detection in Untrimmed Videos, Min Yang, Huan Gao, Ping Guo, Limin Wang
Can’t Make an Omelette Without Breaking Some Eggs: Plausible Action Anticipation Using Large Video-Language Models, Himangi Mittal, Nakul Agarwal, Shao-Yuan Lo, Kwonjoon Lee
End-to-End Temporal Action Detection with 1B Parameters Across 1000 Frames, Shuming Liu, Chen-Lin Zhang, Chen Zhao, Bernard Ghanem
RMem: Restricted Memory Banks Improve Video Object Segmentation, Junbao Zhou, Ziqi Pang, Yu-Xiong Wang
Low-power Continuous Remote Behavioral Localization with Event Cameras, Friedhelm Hamann, Suman Ghosh, Ignacio Juarez Martinez, Tom Hart, Alex Kacelnik, Guillermo Gallego
Action Scene Graphs for Long-Form Understanding of Egocentric Videos, Ivan Rodin, Antonino Fumari, Kyle Min, Subarna Tripathi, Giovanni Maria Farinella
Uncertainty-aware Action Decoupling Transformer for Action Anticipation, Hongji Guo, Nakul Agarwal, Shao-Yuan Lo, Kwonjoon Lee, Qiang Ji
Error Detection in Egocentric Procedural Task Videos, Shih-Po Lee, Zijia Lu, Zekun Zhang, Minh Hoai, Ehsan Elhamifar
Learning to Predict Activity Progress by Self-Supervised Video Alignment, Gerard Donahue, Ehsan Elhamifar
MaskCLaR: Attention-Guided Contrastive Learning for Robust Action Representation Learning, Mohamed Abdelfattah, Mariam Hassan, Alexandre Alahi
Align Before Adapt: Leveraging Entity-to-Region Alignments for Generalizable Video Action Recognition, Yifei Chen, Dapeng Chen, Ruijin Liu, Sai Zhou, Wenyuan Xue, Wei Peng
DIBS: Enhancing Dense Video Captioning with Unlabeled Videos via Pseudo Boundary Enrichment and Online Refinement, Hao Wu, Huabin Liu, Yu Qiao, Xiaojun Sun
 Bridging the Gap: A Unified Video Comprehension Framework for Moment Retrieval and Highlight Detection, Yicheng Xiao, Zhiyu Zhao, Yong Liu, Sai Zhi, Jinyu Yang, Xiu Li
Test-Time Zero-Shot Temporal Action Localization, Benedetta Liberati, Alessandro Conti, Paolo Rota, Yiming Wang, Elisa Ricci
Selective Interpretable and Motion Consistent Privacy Attribute Obfuscation for Action Recognition, Filipp Ilic, He Zhao, Thomas Pock, Richard P. Wildes
Step Differences in Instructional Video, Tushar Nagarajan, Lorenzo Torresani
Compositional Video Understanding with Spatiotemporal Structure-based Transformers, Hoyeoung Yun, Jinwoo Ahn, Minseo Kim, Eun-Sol Kim
Temporally Consistent Unbalanced Optimal Transport for Unsupervised Action Segmentation, Ming Xu, Stephen Gould
FineParser: A Fine-grained Spatio-temporal Action Parser for Human-centric Action Quality Assessment, Jinglin Xu, Sibo Yin, Guochao Zhao, Zishuo Wang, Yuxin Peng
Part-aware Unified Representation of Language and Skeleton for Zero-shot Action Recognition, Anqi Zhu, Qiuhong Ke, Mingming Gong, James Bailey
vid-TLD: Training Free Token Merging for Light-weight Video Transformer, Joonmyung Choi, Sanghyeok Lee, Jaewon Chu, Minhyuk Choi, Hyunwoo J. Kim
CPR-Coach: Recognizing Composite Error Actions based on Single-class Training, Shunli Wang, Shuaibing Wang, Dingkang Yang, Mingchong Li, Haopeng Kuang, Xiao Zhao, Liuzhen Su, Peng Zhai, Lihua Zhang
Uncovering What Why and How: A Comprehensive Benchmark for Causation Understanding of Video Anomaly, Hang Du, Sicheng Zhang, Binzhu Xu, Guoshun Nan, Jiayang Zhang, Junrui Xu, Hangyu Liu, Sичong Leng, Jiangming Liu, Hehe Fan, Dajiu Huang, Jing Feng, Lini Chen, Can Zhang, Xuhuan Li, Hao Zhang, Jianhang Chen, Qimei Cui, Xiaofeng Tao
Detours for Navigating Instructional Videos, Kumar Ashutosh, Zhihu Xue, Tushar Nagarajan, Kristen Grauman
Multiscale Video Transformers Meet Bipartite Matching for Efficient Single-stage Action Localization, Joanna Ntinou, Enrique Sanchez, Georgios Tziropoulos
TE-TAD: Towards Full End-to-End Temporal Action Detection via Time-Aligned Coordinate Expression, Ho-Joong Kim, Jung-Ho Hong, Heejo Kong, Seong-Whan Lee
CSTA: CNN-based Spatiotemporal Attention for Video Summarization, Jaewon Son, Jaehun Park, Kwangsu Kim
PeVL: Pose-Enhanced Vision-Language Model for Fine-Grained Human Action Recognition, Haozong Zhang, Mei Chee Leong, Liuyan Li, Weiis Lin
MULDE: Multiscale Log-Density Estimation via Denoising Score Matching for Video Anomaly Detection, Jakub Micorek, Horst Possegger, Dominik Narrhofer, Horst Bischof, Mateusz Kozinski
Language Model Guided Interpretable Video Action Reasoning, Ning Wang, Guangming Zhu, HS Li, Liang Zhang, Syed Afqa Ali Shah, Mohammed Bennamoun
OST: Refining Text Knowledge with Optimal Spatio-Temporal Descriptor for General Video Recognition, Tongjia Chen, Hongshan Yu, Zhengeng Yang, Zechuan Li, Wei Sun, Chen Chen
Text Prompt with Normality Guidance for Weakly Supervised Video Anomaly Detection, Zhiwei Yang, Jing Liu, Peng Wu
VideoGrounding-DINO: Towards Open-Vocabulary Spatio-Temporal Video Grounding, Syed Talal Wasim, Mumtasm Naseer, Salman Khan, Ming-Hsuan Yang, Fahad Shahbaz Khan
Unsupervised Video Domain Adaptation with Masked Pre-Training and Collaborative Self-Training, Arun Reddy, William Paul, Corban Rivera, Ketul Shah, Celsso de Melo, Rama Chellappa
SnAG: Scalable and Accurate Video Grounding, Fangzhou Mu, Sicheng Mo, Yin Li
Learning Correlation Structures for Vision Transformers, Manjin Kim, Paul Hongsuck Seo, Cordelia Schmid, Minsu Cho
Matching Anything by Segmenting Anything, Suyuan Li, Lei Ke, Martin Danelljan, Luigi Piccinelli, Mattia Segu, Luc Van Gool, Fischer Yu
3D Feature Tracking via Event Camera, Siqi Li, Zhikuan Zhou, Zhou Xue, Yipeng Li, Shaoyi Du, Yue Gao
Frequency Decoupling for Motion Magnification via Multi-Level Isomorphic Architecture, Fei Wang, Dan Guo, Kun Li, Zhun Zhong, Meng Wang
Towards Generalizable Multi-Object Tracking, Zheng Qin, Le Wang, Sanping Zhou, Panpan Fu, Gang Hua, Wei Tang
SocialCircle: Learning the Angle-based Social Interaction Representation for Pedestrian Trajectory Prediction, Conghao
Wong, Beihao Xia, Ziqian Zou, Yulong Wang, Xinge You
Self-Supervised Multi-Object Tracking with Path Consistency, Zijia Lu, Bing Shuai, Yanbei Chen, Zhenlin Xu, Davide Modolo
UnSupFlow: Unsupervised Optical Flow Guided by Segment Anything Model, Shuai Yuan, Lei Luo, Zhao Hui, Can Pu, Xiaoyu Xiang, Rakesh Ranjan, Denis Demandolx
RTracker: Recoverable Tracking via PN Tree Structured Memory, Yuqing Huang, Xin Li, Zikun Zhou, Yoawei Wang, Zhenyu He, Ming-Hsuan Yang
ARTrackV2: Prompting Autoregressive Tracker Where to Look and How to Describe, Yifan Bai, Zeyang Zhao, Yihong Gong, Xing Wei
Endow SAM with Keen Eyes: Temporal-spatial Prompt Learning for Video Camouflaged Object Detection, Wenjun Hui, Zhenfeng Zhu, Shuai Zheng, Yao Zhao
MemFlow: Optical Flow Estimation and Prediction with Memory, Qiaolee Dong, Yanwei Fu
OneTracker: Unifying Visual Object Tracking with Foundation Models and Efficient Tuning, Lingyi Hong, Shilin Yan, Renrui Zhang, Wanyun Li, Xinyu Zhou, Pinxue Guo, Kaixun Jiang, Yiling Chen, Jinglin Li, Zhaoyu Chen, Wengiang Zhang
Learned Trajectory Embedding for Subspace Clustering, Yaroslava Lochman, Carl Olsson, Christopher Zach
PNeRV: Enhancing Spatial Consistency via Pyramidal Neural Representation for Videos, Qi Zhao, M. Salman Asif, Zhan Ma
DiffusionTrack: Point Set Diffusion Model for Visual Object Tracking, Fei Xie, Zhongdaowang, Chao Ma
Sparse Global Matching for Video Frame Interpolation with Large Motion, Chunxu Liu, Guozhen Zhang, Rui Zhao, Limin Wang
iKUN: Speak to Trackers without Retraining, Yunhao Du, Cheng Lei, Zicheng Zhao, Fei Su
NetTrack: Tracking Highly Dynamic Objects with a Net, Guangze Zheng, Shijie Lin, Haobo Zuo, Changhong Fu, Jia Pan
FlowDiffuser: Advancing Optical Flow Estimation with Diffusion Models, Ao Luo, Xin Li, Fan Yang, Jiangyu Liu, Haoqiang Fan, Shuaicheng Liu
Video Harmonization with Triple Spatio-Temporal Variation Patterns, Zonghui Guo, Xinyu Han, Jie Zhang, Shiguang Shan, Haiyong Zheng
Dense Optical Tracking: Connecting the Dots, Guillaume Le Moing, Jean Ponce, Cordelia Schmid
Efficient Meshflow and Optical Flow Estimation from Event Cameras, Xinglong Luo, Ao Luo, Zhengning Wang, Chunyong Lin, Bing Zeng, Shuaicheng Liu
Context-Aware Integration of Language and Visual References for Natural Language Tracking, Yanyan Shao, Shuting He, Qi Ye, Yuchao Feng, Wenhan Luo, Jiming Chen
Depth-aware Test-Time Training for Zero-shot Video Object Segmentation, Weihuang Liu, Xi Shen, Haolun Li, Xiuli Bi, Bo Liu, Chi-Man Pun, Xiaodong Cun
Weakly Supervised Video Individual Counting, Xinyan Liu, Guorong Li, Yuankai Qi, Ziheng Yan, Zhenjun Han, Anton van den Hengel, Ming-Hsuan Yang, Qingming Huang
Dual Prototype Attention for Unsupervised Video Object Segmentation, Suwan Cho, Minhyeok Lee, Seunghoon Lee, Dogyoon Lee, Heeseung Choi, Ig-Jae Kim, Sangyoun Lee
HIPTrack: Visual Tracking with Historical Prompts, Wenrui Cai, Qingjie Liu, Yunhong Wang
FlowTrack: Revisiting Optical Flow for Long-Range Dense Tracking, Seokju Cho, Jiahui Huang, Seungryong Kim, Joon-Young Lee
Implicit Motion Function, Yue Gao, Jiahao Li, Lei Chu, Yan Lu
DeconfuseTrack: Dealing with Confusion for Multi-Object Tracking, Cheng Huang, Shoudong Han, Mengyu He, Wenbo Zheng, Yuhaowei Notes:

## 19:00 - 21:00 Reception & Musical Performances
(Summit Flex Hall ABC)

### Notes:
Friday, June 21

8:00 - 14:00  Registration / Badge Pickup (Summit Lobby)
7:00 - 17:00  Prayer or Quiet Room (Summit 341-adjacent and Summit 441-adjacent)
7:00 - 17:00  Prayer or Quiet Room (Summit 336)
8:00 - 9:30  Breakfast (Summit ExHall 1-2)
9:00 - 10:00  EXPO TRACK - Ece Kamar, VP, Managing Director of AI Frontiers at Microsoft Research (Arch 4F)
9:00 - 10:30  ORALS 5A: Datasets and Evaluation (Summit Ballroom)

1 Deep Generative Model based Rate-Distortion for Image Downscaling Assessment, Yuanbang Liang, Bhavesh Garg, Paul Rosin, Yipeng Qin

2 360x+: A Panoptic Multi-modal Scene Understanding Dataset, Hao Chen, Yuqi Hou, Chenyuan Qu, Irene Testini, Xiaohan Hong, Jianbo Jiao


4 Rich Human Feedback for Text-to-Image Generation, Youwei Lia, Junfeng He, Gang Li, Peizhe Liu, Liangyi Yan, Xingchen Yang, Zhewang Ye, Shengxin Cindy Zha, Chen Zhao, Zhiwei Zhao, Zhihan Zhu, Jeff Zhuo, Pablo Arbelaez, Gedas Bertasius, Dimma Damen, Jakob Engel, Giovanni Maria Farinella, Antonino Furnari, Bernard Ghanem, Judy Hoffman, C.W. Jawahar, Richard Newcombe, Hyun Soo Park, James M. Rehg, Yoichi Sato, Manolis Savva, Jianbo Shi, Mike Zheng Shou, Michael Wray

5 Human Feedback for Text-to-Image Generation, Youwei Lia, Junfeng He, Gang Li, Peizhe Liu, Liangyi Yan, Xingchen Yang, Zhewang Ye, Shengxin Cindy Zha, Chen Zhao, Zhiwei Zhao, Zhihan Zhu, Jeff Zhuo, Pablo Arbelaez, Gedas Bertasius, Dimma Damen, Jakob Engel, Giovanni Maria Farinella, Antonino Furnari, Bernard Ghanem, Judy Hoffman, C.W. Jawahar, Richard Newcombe, Hyun Soo Park, James M. Rehg, Yoichi Sato, Manolis Savva, Jianbo Shi, Mike Zheng Shou, Michael Wray

9:00 - 10:30  ORALS 5B: 3D from Multiview and Sensors (Summit Flex Hall AB)

1 Grounding and Enhancing Grid-based Models for Neural Fields, Zelin Zhao, Fenglei Fan, Wenlong Liao, Junchi Yan, Taorong Leng

2 NeRF-Hugs: Improved Neural Radiance Fields in Non-static Scenes Using Heuristics-Guided Segmentation, Jiahao Chen, Yipeng Qin, Lingjie Li, Jiangbo Liu, Quanbin Li, Min Li, Weitao Guo, Wei-Ming Sun, Bo-Wen Tai, Hongyi Wu, Chao Xu, Xinyu Zeng, Shaojun Zhou

3 Mip-Splatting: Alias-free 3D Gaussian Splatting, Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, Andreas Geiger

4 PixelSplat: 3D Gaussian Splat from Image Pairs for Scalable Generalizable 3D Reconstruction, David Charatan, Szhe Lester Li, Andrea Tagliasacchi, Vincent Sitzmann

5 Learning to Produce Semi-dense Correspondences for Visual Localization, Khaung Truong Giang, Soohwan Song, Sungho Jo

9:00 - 10:30  ORALS 5C: Low-Shot, Self-Supervised, Semi-Supervised Learning (Summit Flex Hall C)

1 CroSel: Cross Selection of Confident Pseudo Labels for Partial-Label Learning, Shiyu Tian, Hongxin Wei, Yiqun Wang, Lei Feng

2 MLP Can Be A Good Transformer Learner, Shihan Lin, Pumeng Lyu, Dongrui Liu, Tao Tang, Xiaodan Liang, Andy Song, Xiaojun Chang

3 From SAM to CAMs: Exploring Segment Anything Model for Weakly Supervised Semantic Segmentation Hyeokjun Kweon, Kuk-jin Yoon

4 LTGC: Long-tail recognition via leveraging LLMs-driven Generated Content, Qihao Zhao, Yalan Dai, Hao Li, Wei Hu, Fan Zhang, Jun Liu

5 Improving Semantic Correspondence with Viewpoint-Guided Spherical Maps, Octave Mariotti, Oisin Mac Aodha, Hakan Bilen

10:00 - 10:30  Poster Setup (Arch 4E)

10:00 - 12:00  Poster Session 5 & Exhibit Hall (Arch 4A-E)

1 ToMo: Towards Text-Driven 3D Styling for Multi-Object Meshes, Xuying Zhang, Bo-Wen Yin, Yuming Chen, Zheng Lin, Yunheng Li, Qibin Hou, Ming-Ming Cheng

2 Event-based Structure-from-Orbit, Ethan Elms, Yasir Latif, Tae Ha Park, Tat-Jun Chun

3 Towards Large-scale 3D Representation Learning with Multi-dataset Point Prompt Learning, Xiaoyang Wu, Zhuotao Tian, Xin Wen, Bohao Peng, Xiului Li, Kaicheng Yu, Hengshuang Zhao

4 LidaRF: Delving into Lidar for Neural Radiance Field on Street Scenes, Shanlin Sun, Bingbing Zhuang, Ziyu Jiang, Buyu Liu, Xiaohui Xie, Manmohan Chandraker

5 Instantaneous Perception of Moving Objects in 3D, Di Liu, Bingbing Zhuang, Dimitris N. Metaxas, Manmohan Chandraker

6 Implicit Event-RGBD Neural SLAM, Delin Qu, Chi Yan, Dong Wang, Je Yi, Qizhi Chen, Dan Xu, Yiting Zhang, Bin Zhao, Xuelong Li

7 GS-SLAM: Dense Visual SLAM with 3D Gaussian Splatting, Chi Yan, Delin Qu, Dan Xu, Bin Zhao, Zhigang Wang, Dong Wang, Xuelong Li

8 Learning Instance-Aware Correspondences for Robust Multi-Instance Point Cloud Registration in Cluttered Scenes, Zhiyuan Yu, Zheng Qin, Lintao Zheng, Kai Xu

9 MeshGPT: Generating Triangle Meshes with Decoder-Only Transformers, Yawar Siddiqui, Antonio Allegro, Alexey Artemov, Tatiana Tommasi, Daniele Sirigatti, Vladislav Rosov, Angela Dai, Matthias Nießner

10 Multi-Session SLAM with Differentiable Wide-Baseline Pose Optimization, Lahav Lipson, Jia Deng

11 SHINOBI: Shape and Illumination using Neural Object Decomposition via BRDF Optimization In-the-wild, Andreas Engelhardt, Amit Raj, Mark Boss, Yunzhi Zhang, Abhishek Kar, Yuanzheng Li, Deqing Sun, Ricardo Martin Brual, Jonathan T. Barron, Hendrik P. A. Lensch, Varun Jampani

12 HybridNeRF: Efficient Neural Rendering via Adaptive Volumetric Surfaces, Haithem Turki, Asad Agrawal, Samuel Rota Bulò, Lorenzo Porzi, Peter Kortschieder, Deva Ramanan, Michael Zollhöfer, Christian Richard

13 PLGSLAM: Progressive Neural Scene Representation with Local to Global Bundle Adjustment, Tianchen Deng, Guole Shen, Tong Qian, Jianyu Wang, Wentao Zhao, Jingchuan Wang, Danwei Wang, Weidong Chen

14 Gear-NeRF: Free-Viewpoint Rendering and Tracking with Motion-aware Spatio-Temporal Sampling, Xinhang Liu, Yu-Wing Tai, Chi-Keung Tang, Pedro Míraldo, Suhas Lohit, Moitreya Chatterjee


16 HyperSDFusion: Bridging Hierarchical Structures in Language and Geometry for Enhanced 3D Text2Shape Generation, Zhiying Leng, Tolga Birdal, Xiaohui Liang, Federico Tombari

18 Animatable Gaussians: Learning Pose-dependent Gaussian Maps for High-fidelity Human Avatar Modeling, Zhe Li, Zerong Zhang, Lizhen Wang, Yebin Liu

19 Global Latent Neural Rendering, Thomas Tanay, Matteo Maggioni

20 HiFi4G: High-Fidelity Human Performance Rendering via Compact Gaussian Splatting, Yuheng Jiang, Zhehao Shen, Penghao Wang, Zhuo Su, Yu Hong, Yingliang Zhang, Jingyi Yu, Lan Xu

21 LoS: Local Structure-Guided Stereo Matching, Kunhong Li, Longguang Wang, Ye Zhang, Kaixen Xue, Shunbo Zhu, Yulan Guo


23 Masked Spatial Propagation Network for Sparsity-Adaptive Depth Refinement, Jinyoung Jun, Jae-Han Lee, Chang-Su Kim

24 CausalPC: Improving the Robustness of Point Cloud Classification by Causal Effect Identification, Yuanmin Huang, Mi Zhang, Daizong Ding, Erling Jiang, Zhaoxiang Wang, Min Yang

25 RoMa: Robust Dense Feature Matching, Johan Edstedt, Qiuy Sun, Georg Bökman, Mårten Wadenbäck, Michael Felsberg

26 MVHumanNet: A Large-scale Dataset of Multi-view Daily Dressing Human Captures, Zhangyang Xiong, Chenghong Li, Kunkun Liu, Hongjie Liao, Jiangqiao Hu, Junyi Zhu, Shuliang Ning, Lingteng Qiu, Chongjie Wang, Shijie Wang, Shuguan Cui, Xiaoguang Han


28 RegionPLC: Regional Point-Language Contrastive Learning for Open-World 3D Scene Understanding, Jihan Yang, Runyu Ding, Weipeng Deng, Zhe Wang, Xiaojuan Qi

29 NeLF-Pro: Neural Light Field Probes for Multi-Scalar Novel View Synthesis, Zinuo You, Andreas Geiger, Anpei Chen


31 FAR: Flexible Accurate and Robust 6DOF Relative Camera Pose Estimation, Chris Rockwell, Niles Kulkarni, Linyi Jin, Jeong Joon Park, Justin Johnson, David F. Fouhey

32 OmniGlue: Generalizable Feature Matching with Foundation Model Guidance, Hanwen Jiang, Arjun Karpur, Bingyi Cao, Qixing Huang, André Araujo

33 GART: Gaussian Articulated Template Models, Jiahui Lei, Yufu Wang, Georgios Pavlakos, Lingjie Liu, Kostas Estathopoulos, Jiahao Chen, Yipeng Qin, Lingjie Liu, Jiangbo Lu, Guanbin Li

34 PROGRAM GUIDE

35 S-DyRF: Reference-Based Stylized Radiance Fields for Dynamic Scenes, Xingyi Li, Zhiguo Cao, Yizheng Wu, Kewei Wang, Ke Xian, Zhe Wang, Guosheng Lin, Hanspeter Pfister

36 BEVNeXt: Reviving Dense BEV Frameworks for 3D Object Detection, Zhenxin Li, Shiyi Lan, Jose M. Alvarez, Zuxuan Wu

37 Bi-SCC: Geometric-Semantic Bidirectional Fusion for Camera-based 3D Semantic Scene Completion, Yujie Xue, Ruihui Li, Fan Wu, Zhuo Tang, Kenli Li, Mingxing Duan

38 Learning to Select Views for Efficient Multi-View Understanding, Yinzhong Hou, Stephen Gould, Liang Zheng

39 Outdoor Scene Extrapolation with Hierarchical Generative Model, Dongsu Zhang, Francis Williams, Zan Gojcic, Karsten Kreis, Sanja Fidler, Young Min Kim, Amlan Kar

40 Spectrum AUC Difference (SAUCD): Human-aligned 3D Shape Evaluation, Tianyu Luan, Zhong Li, Lele Chen, Xuan Gong, Lichang Chen, Yi Xu, Junsong Yuan

41 Federated Online Adaptation for Deep Stereo, Matteo Poggi, Fabio Tosi

42 Instruct 4D-to-4D: Editing 4D Scenes as Pseudo-3D Scenes Using 2D Diffusion, Linzhan Mou, Jun-Kun Chen, Yu-Xiong Wang

43 Real-time Acquisition and Reconstruction of Dynamic Volumes with Neural Structured Illumination, Yixun Zeng, Zoubin Bi, Mingrui Yin, Xiang Feng, Kun Zhou, Hongzhi Wu

44 Unifying Correspondence Pose and NeRF for Generalized Pose-Free Novel View Synthesis, Sungkhun Hong, Jaewoo Jung, Heeseong Shin, Jiaolong Yang, Seungeyong Kim, Chong Luo

45 GoMVS: Geometrically Consistent Cost Aggregation for Multi-View Stereo, Jiwu Li, Rui Li, Haofei Xu, Wenxun Zhao, Yu Zhu, Jinqu Sun, Yanning Zhang

46 MESA: Matching Everything by Segmenting Anything, Yesheng Zhang, Xu Zhao

47 OmniSFDF: Scene Reconstruction using Omnidirectional Signed Distance Functions and Adaptive Bincotrees, Hakyeong Kim, Andreas Meuleman, Hyeonjoo Kim, James Tompkin, Min H. Kim

48 MirageRoom: 3D Scene Segmentation with 2D Pre-trained Models by Mirage Projection, Haowen Sun, Yueqi Duan, Juncheng Yan, Yifan Liu, Jiwen Lu

49 Robust Synthetic-to-Real Transfer for Stereo Matching, Jiawei Zhang, Jihe Li, Lei Huang, Xiaohan Yu, Lin Gu, Jin Zheng, Xiao Bai
Symphonize 3D Semantic Scene Completion with Contextual Instance Queries, Haoyi Jiang, Tianheng Cheng, Naiyu Gao, Haoyang Zhang, Tianwei Lin, Wenyu Liu, Xingang Wang

Differentiable Neural Surface Refinement for Modeling Transparent Objects, Weijian Deng, Dylan Campbell, Chunyi Sun, Shubham Kanitkar, Matthew E. Shaffer, Stephen Gould

DeMatch: Deep Decomposition of Motion Field for Two-View Correspondence Learning, Shihua Zhang, Zizhao Li, Yuan Gao, Jiayi Ma

Is Vanilla MLP in Neural Radiance Field Enough for Few-shot View Synthesis?, Hanxin Zhu, Tianyu He, Xin Li, Bingchen Li, Zhibo Chen

GaussianAvatars: Photorealistic Head Avatars with Rigged 3D Gaussians, Shenhao Qian, Tobias Kirschstein, Liam Schoneveld, Davide Davoli, Simon Giebenhain, Matthias Nießner

4D Gaussian Splatting for Real-Time Dynamic Scene Rendering, Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, Xingang Wang

How Far Can We Compress Instant-NGP-Based NeRF?, Yihang Chen, Qianyi Wu, Mehrtaş Harandi, Jianfei Cai

Deformable 3D Gaussians for High-Fidelity Monocular Dynamic Scene Reconstruction, Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, Xiaogang Jin

Learning with Unreliability: Fast Few-shot Voxel Radiance Fields with Relative Geometric Consistency, Yingjie Xu, Bangzhen Li, Hao Tang, Bailin Deng, Shengfeng He

NTO3D: Neural Target Object 3D Reconstruction with Segment Anything, Xiaobao Wei, Renrui Zhang, Jiarui Wu, Jiangbing Liu, Ming Lu, Yandong Guo, Shanghang Zhang

Loopy-SLAM: Dense Neural SLAM with Loop Closures, Lorenzo Liso, Erik Sandstrom, Vladimir Yagay, Luca Van Gool, Martin R. Oswald

BSNet: Box-Supervised Simulation-assisted Mean Teacher for 3D Instance Segmentation, Jiahao Lu, Jiaheng Ding, Tianshu Zhang

ExtraNeRF: Visibility-Aware View Extrapolation of Neural Radiance Fields with Diffusion Models, Meng-Li Shih, Wei-Chiu Ma, Lorenzo Boyice, Aleksander Holynski, Forrester Cole, Brian Curless, Janne Kontkanen

Alpha Invariance: On Inverse Scaling Between Distance and Volume Density in Neural Radiance Fields, Joshua Ahn, Haochen Wang, Raymond A. Yeh, Greg Shakharovich

SpatialTracker: Tracking Any 2D Pixels in 3D Space, Yuxi Xiao, Qianqian Wang, Shangzhan Zhang, Nan Xue, Sida Peng, Yujun Shen, Xiaowei Zhou

GauHuman: Articulated Gaussian Splatting from Monocular Human Videos, Shoukun Hu, Tao Hu, Ziwei Liu

IPoD: Implicit Field Learning with Point Diffusion for Generalizable 3D Object Reconstruction from Single RGB-D Images, Yushuang Wu, Luyue Shi, Junhao Cai, Weihao Yuan, Lingteng Qiu, Zilong Dong, Liebong Bo, Shuguang Cui, Xiaoguang Han

GOV-NeSF: Generalizable Open-Vocabulary Neural Semantic Fields, Yunsong Wang, Hanlin Chen, Gin Hee Lee

LASA: Instance Reconstruction from Real Scans using A Large-scale Aligned Shape Annotation Dataset, Haolun Liu, Chongjie Ye, Yinyu Nie, Yingfan He, Xiaoguang Han

GenZi: Zero-Shot 3D Human-Scene Interaction Generation, Lei Li, Angela Dai

MVCPS-NeuS: Multi-view Constrained Photometric Stereo for Neural Surface Reconstruction, Hiroaki Santo, Fumio Okura, Yasuyuki Matsushita

DVMNet: Computing Relative Pose for Unseen Objects Beyond Hypotheses, Chen Zhao, Tong Zhang, Zheng Dang, Mathieu Salzmann

Motion2VecSets: 4D Latent Vector Set Diffusion for Non-rigid Shape Reconstruction and Tracking, Wei Cao, Chang Luo, Biao Zhang, Matthias Nießner, Jiapeng Tang

DiffScene: Denoising Diffusion Models for Generative Indoor Scene Synthesis, Jiapeng Tang, Yuyu Nie, Lev Markhasin, Angela Dai, Justus Thies, Matthias Nießner

Test-Time Adaptation for Depth Completion, HyoYoung Seok Park, Anjali Gupta, Alex Wong

Global and Hierarchical Geometry Consistency Priors for Few-shot NeRFs in Indoor Scenes, Xiaotian Sun, Qingshan Xu, Xinjie Yang, Yu Zang, Cheng Wang

KPRED: Exploiting Semantic Keypoints for Joint 3D Shape Retrieval and Deformation, Ruida Zhang, Chenyangguang Zhang, Yan Di, Fabian Manhardt, Xinyu Liu, Federico Tombari, Xiangyang Ji

Unsigned Orthogonal Distance Fields: An Accurate Neural Implicit Representation for Diverse 3D Shapes, Yujie Lu, Long Wan, Nanyi Ding, Yulong Wang, Shuhan Shen, Shen Cai, Lin Gao

DISR-NeRF: Diffusion-Guided View-Consistent Super-Resolution NeRF, Jie Long Lee, Chen Li, Gini Hee Lee

BANF: Band-Limited Neural Fields for Levels of Detail Reconstruction, Akhmedkhan Shabanov, Shrisudhan Govindarajan, Cody Reading, Lily Goli, Daniel Reiben, Kwang Moo Yi, Andrea Tagliasacchi

SuperNormal: Neural Surface Reconstruction via Multi-View Normal Integration, Xu Cao, Takafumi Taketomi

ADFactory: An Effective Framework for Generalizing Optical Flow with NeRF, Han Ling, Quansun Sun, Yinghui Sun, Xian Xu, Xinfeng Li

DrHair: Reconstructing Scalp-Connected Hair Strands without Pre-Training via Differentiable Rendering of Line Segments, Yusuke Takimoto, Hikari Takehara, Hiroyuki Sato, Zhaohu Bo, Zheng

Omniseg3D: Omniversal 3D Segmentation via Hierarchical Contrastive Learning, Haiyai Ying, Yixuan Yin, Jinhzi Zhang, Fan Wang, Tao Yu, Ruqi Huang, Lu Fang

Visual Programming for Zero-shot Open-Vocabulary 3D Visual Grounding, Zhihao Yuan, Jinke Ren, Chun-Mei Feng, Hengshuang Zhao, Shuguang Cui, Zhen Li


Edge-Aware 3D Instance Segmentation Network with Intelligent Semantic Prior, Wonseok Roh, Hwanhee Jung, Giljoo Nam, Jinsop Yeom, Hyunjie Park, Sang Ho Yoon, Sangpil Kim

Scaffold-GS: Structured 3D Gaussians for View-Adaptive Rendering, Tao Lu, Mulin Yu, Liming Xu, Yuanbo Xiangli, Limin Wang, Dahua Lin, Bo Dai

Map-Relative Pose Regression for Visual Re-Localization, Shuai Chen, Tommaso Cavalli, Victor Adrian Prisacariu, Eric Brachmann

3DGStream: On-the-Fly Training of 3D Gaussians for Efficient Streaming of Photo-Realistic Free-Viewpoint Videos, Jiakai Sun, Han Jiao, Guangyuan Li, Zhanjie Zhang, Lei Zhao, Wei Xing

Revisiting Global Translation Estimation with Feature Tracks, Peilin Tao, Hainan Cui, Mengdi Rong, Shuhan Shen

DUS3IR: Geometric 3D Vision Made Easy, Shuzhe Wang, Vincent Leroy, Yohann Cabon, Boris Chidlovskii, Jerome Revaud

Robust Depth Enhancement via Polarization Prompt Fusion Tuning, Kei Ikemura, Yiming Huang, Felix Heide, Zhaoxiang Zhang, QiFeng Chen, Chenyang Lei

StraightPCF: Straight Point Cloud Filtering, Dasith de Silva Edirimuni, Xuequan Lu, Gang Li, Lei Wei, Antonio Robles-Kelly, Hongdong Li


NeRF Director: Revisiting View Selection in Neural Volume Rendering, Wenhui Xiao, Rodrigo Santa Cruz, David Ahmedt-Aristizabal, Olivier Salvado, Clinton Fookes, Leandro Lebrat

Learning Intra-view and Cross-view Geometric Knowledge for Stereo Matching, Rui Gong, Weida Liu, Zaiweng Gu, Xufei Yang, Jun Cheng

Sherpa3D: Boosting High-Fidelity Text-to-3D Generation via Coarse 3D Prior, Fangfu Liu, Diankun Wu, Yi Wei, Yongming Rao, Yueqi Duan

DNGaussian: Optmizing Sparse-View 3D Gaussian Radiance Fields with Global-Local Depth Normalization, Jiahe Li, Jiawei Sun, Junhao Cai, Tong Zhang, Zheng Dang, Mathieu Salzmann
120 COLMAP-Free 3D Gaussian Splatting, **Yang Fu, Sifei Liu, Amey Kulkarni, Jan Kautz, Alexei A. Efros, Xiaolong Wang**

121 GSNerF: Generalizable Semantic Neural Radiance Fields with Enhanced 3D Scene Understanding, **Zi-Ting Chou, Sheng-Yu Huang, I-Jieh Liu, Yu-Chiang Frank Wang**

122 Extend Your Own Correspondences: Unsupervised Distant Point Cloud Registration by Progressive Distance Extension, **Quan Liu, Hongzi Zhu, Zhenxi Wang, Yusong Zhou, Shan Chang, Minyi Guo**

123 Fully Geometric Panoramic Localization, **Junho Kim, Jiwon Jeong, Young Min Kim**

124 Multiway Point Cloud Mosaicking with Diffusion and Global Optimization, **Shengze Jin, Iro Armeni, Marc Pollefeys, Daniel Barth**

125 Mip-Splatting: Alias-free 3D Gaussian Splatting, **Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, Andreas Geiger**

126 Generative 3D Part Assembly via Part-Whole-Hierarchy Message Passing, **Bian Du, Xiang Gao, Wei Hu, Renjie Liao**

127 Total-Decom: Decomposed 3D Scene Reconstruction with Minimal Interaction, **Xiaoyang Lyu, Chirui Chang, Peng Dai, Yang-Tian Sun, Xiaojuan Qi**

128 Absolute Pose from One or Two Scaled and Oriented Features, **Jonathan Ventura, Zuzana Kukelova, Torsten Sattler, Daniel Barath**

129 DGC-GNN: Leveraging Geometry and Color Cues for Visual Descriptor-Free 2D-3D Matching, **Shuzhe Wang, Juho Kannala, Daniel Barth**

130 Entity-NeRF: Detecting and Removing Moving Entities in Urban Scenes, **Takashi Otonari, Satoshi Ikehata, Kiyoharu Aizawa**

131 GaussianEditor: Editing 3D Georgians Delicately with Text Instructions, **Junjie Wang, Jieming Fan, Xiaopeng Zhang, Lingsi Xie, Qi Tian**

132 The Moey You See in 2D the more You Perceive in 3D, **Xingyuan Han, Zelin Gao, Anjoo Kanazawa, Shubham Goel, Yossi Gandelsman**

133 Multi-Scale 3D Gaussian Splatting for Aliased Rendering, **Zhizen Yan, Feng Fei Leu, Yu Chen, Jim Hoon Lee**

134 Practical Measurements of Translucency Materials with Inter-Pixel Translucency Prior, **Zhenyu Chen, Jie Guo, Shuichang Lai, Ruoyu Fu, Mengxun Kong, Chen Wang, Hongyu Sun, Zhebin Zhang, Chen Li, Yanwen Guo**

135 OneFormer3D: One Transformer for Unified Point Cloud Segmentation, **Maxim Kolodizhnyi, Anna Vorontsova, Anton Konushin, Danila Rukhovich**

136 General Point Model Pretraining with Autoencoding and Autoregressive, **Zhe Li, Zhangyang Gao, Cheng Tan, Bocheng Ren, Laurence T. Yang, Stan Z. Li**

137 MorpheUS: Neural Dynamic 360° Surface Reconstruction from Monocular RGB-D Video, **Hengyi Wang, Jingwen Wang, Lourdes Agapito**

138 pixelSplat: 3D Gaussian Splits from Image Pairs for Scalable Generalizable 3D Reconstruction, **David Charatan, Szhe Lester Li, Andrea Tagliasacchi, Vincent Sitzmann**

139 Object Dynamics Modeling with Hierarchical Point Cloud-based Representations, **Changho Kim, Li Fuxin**

140 Neural Refinement for Absolute Pose Regression with Feature Fusion, **Shuai Chen, Yash Bhagat, Xinghui Li, Jia-Wang Bian, Kejie Li, Zirui Wang, Victor Adrian Prisacariu**

141 Gaussian Shadow Casting for Neural Characters, **Luis Bolanos, Shih-Yang Su, Helge Rhodin**

142 PAPR in Motion: Seamless Point-level 3D Scene Interpolation, **Chicheng Peng, Yanshu Zhang, Ke Li**

143 ShapeMatcher: Self-Supervised Joint Shape Canonicalization Segmentation Retrieval and Deformation, **Yan Di, Chenyangguang Zhang, Chaowei Wang, Ruida Zhang, Guangyao Zhai, Yanjun Li, Bowen Fu, Xiangyang Ji, Shan Gao**

144 XScale-NVS: Cross-Scale Novel View Synthesis with Hash Featureized Manifold, **Guangyu Wang, Jinzhong Zhan, Fan Wang, Ruiqi Huang, Lu Fang**

145 Instance-Adaptive and Geometric-Aware Keypoint Learning for Category-Level 6D Object Pose Estimation, **Xiao Lin, Wenfei Yang, Yuan Gao, Tianzhu Zhang**

146 RepKPU: Point Cloud Upsampling with Kernel Point Representation and Deformation, **Yi Rong, Haoran Zhou, Kang Xia, Cheng Mei, Jiahao Wang, Tong Lu**

147 ColorPCR: Color Point Cloud Registration with Multi-Stage Geometric-Color Fusion, **Junchao Mu, Lin Bie, Shaoyi Du, Yue Gao**

148 Consistent3D: 3D-Consistent 2D Diffusion for High-Fidelity Scene Editing, **Jun-Kun Chen, Samuel Rota Bulò, Norman Müller, Lorenzo Porzi, Peter Kuntschieder, Yu-Xiong Wang**

149 SceneTex: High-Quality Texture Synthesis for Indoor Scenes via Diffusion Priorers, **Dave Zhenyu Chen, Haoxuan Li, Hsin-Ying Lee, Sergey Tulyakov, Matthias Nießner**

150 Aerial Lifting: Neural Urban Semantic and Building Instance Lifting from Aerial Imagergy, **Yuqi Zhang, Guanying Chen, Xiaochen Shuguang Cui**

151 Improving Depth Completion via Depth Feature Upsampling, **Yufei Wang, Ge Zhang, Shaoqian Wang, Bo Li, Qi Liu, Le Hui, Yuchao Dai**

152 ZeroRF: Fast Sparse View 360° Reconstruction with Zero Pretraining, **Roxui Shi, Xinyue Wei, Cheng Wang, Hau Su**

153 Multi-Level Neural Scene Graphs for Dynamic Urban Environments, **Tobias Fischer, Lorenzo Porzi, Samuel Rota Bulò, Marc Pollefeys, Peter Kuntschieder**

154 Gaussian-Flow: 4D Reconstruction with Dynamic 3D Gaussian Particle, **Youtian Lin, Zuxiaozhu Dai, Siyu Zhu, Yao Yao**

155 L4D-Track: Language-to-4D Modeling Towards 6-DoF Tracking and Shape Reconstruction in 3D Point Cloud Stream, **Jingtao Sun, Yoonan Wang, Mingtao Feng, Yulan Guo, Ajmal Mian, Mike Zheng Shou**

156 Neural Directional Encoding for Efficient and Accurate View-Dependent Appearance Modeling, **Liwen Wu, Sai Bi, Zexiang Xu, Fujun Luan, Kai Zhang, Ilyan Georgiev, Kalyan Sunkavalli, Ravi Ramamoorthi**

157 SNI-SLAM: Semantic Neural Implicit SLAM, **Siting Zhu, Guangming Wang, Hermann Blum, Jiuming Liu, Liang Song, Marc Pollefeys, Hesheng Wang**

158 Enhancing 3D Object Detection with 2D Detection-Guided Query Anchors, **Haoxuan Li, Pengpeng Liang, Erkang Cheng**

159 SpecNeRF: Gaussian Directional Encoding for Specular Reflections, **Li Ma, Vasu Agrawal, Haithem Turki, Changli Kim, Chen Gao, Pedro Sander, Michael Zollhöfer, Christian Richardt**

160 Correspondence-Free Non-Rigid Point Set Registration Using Unsupervised Clustering Analysis, **Mingyang Zhao, Jiaoe Feng, Mei Sha, Qingshui Xiong, Dong-Ming Yan**

161 GAFusion: Adaptive Fusing LiDAR and Camera with Multiple Guidance for 3D Object Detection, **Xiaotian Li, Baojie Fan, Jiandong Tian, Huijie Fan**

162 3D Neural Edge Reconstruction, **Lei Li, Songyou Peng, Zehao Yu, Chaohui Liu, Hémi Pautrat, Xiaochuan Yin, Marc Pollefeys**

163 AlignMiF: Geometry-Aligned Multimodal Implicit Field for LiDAR-Camera Joint Synthesis, **Tang Tao, Guangming Wang, Yixiong Lao, Peng Chen, Jie Liu, Liang Lin, Kaicheng Yu, Xiaodan Liang**

164 Polarization Wavefront Lidar: Learning Large-Scale Reconstruction from Polarized Wavefronts, **Dominik Scheuble, Chenyang Lei, Seung-Hwan Baek, Mario Bijaik, Felix Heide**

165 A Unified Diffusion Framework for Scene-aware Human Motion Estimation from Sparse Signals, **Jiangnan Tang, Jingya Wang, Kaiyang Ji, Lan Xu, Jingyi Yu, Ye Shi**

166 FaceTalk: Audio-Driven Motion Diffusion for Neural Parametric Head Models, **Shivangi Aneja, Justus Thies, Angela Dai, Matthias Nießner**

167 NeRFCodec: Neural Feature Compression Meets Neural Radiance Fields for Memory-Efficient Scene Representation, **Sicheng Li, Hao Li, Yi Yi, Liao, Lu Yu**

168 Open-Vocabulary 3D Semantic Segmentation with Foundation Models, **Li Jiang, Shaoshuai Shi, Bernt Schiele**

169 GraphDreamer: Compositional 3D Scene Synthesis from Scene Graphs, **Gege Gao, Weiyang Liu, Anpei Chen, Andreas Geiger, Bernhard Schölkopf**

170 OA-CNNs: Omni-Adaptive Sparse CNNs for 3D Semantic Segmentation, **Bohao Peng, Xiaoyang Wu, Li Jiang, Yukang Chen, Hengshuang Zhao, Zhuotao Tian, Jiaya Jia**

171 Efficient Solution of Point-Line Absolute Pose, **Petr Hruby, Timothy Duff, Marc Pollefeys**
CN-RMA: Combined Network with Ray Marching Aggregation for 3D Indoor Object Detection from Multi-view Images, Guanlin Shen, Jingwei Huang, Zhihua Hu, Bin Wang

HUGS: Holistic Urban 3D Scene Understanding via Gaussian Splatting, Hongyou Zhou, Jiahaoh Shao, Lu Xu, Dongfeng Bai, Weichao Qiu, Bingbing Liu, Yue Wang, Andreas Geiger, Yiyi Liao

Benchmarking Implicit Neural Representation and Geometric Rendering in Real-Time RGB-D SLAM, Tongyan Hua, Lin Wang

SplatAM: Splat Track & Map 3D Gaussians for Dense RGB-D SLAM, Nikhil Keetha, Jay Karhade, Krishna Murthy Jatavallabhula, Gengshan Yang, Sebastian Scherer, Deva Ramanan, Jonathan Luiten


TutteNet: Injective 3D Deformations by Composition of 2D Mesh

Deformations, Bo Sun, Thibault Groueix, Chen Song, Qixing Huang, Noam Aigerman

LO-Sampler: An LO Model Guided Volume Sampling for NeRF, Liangchen Li, Juyong Zhang

Text-to-3D using Gaussian Splatting, Zilong Chen, Feng Wang, Yikai Wang, Huaping Liu

TAMM: TriAdapter Multi-Modal Learning for 3D Shape Understanding, Zhihao Zhang, Shengcao Cao, Yu-Xiong Wang

FreGS: 3D Gaussian Splatting with Progressive Frequency Regularization, Jiahui Zhang, Fangneng Zhan, Muyu Xu, Shijian Lu, Eric Xing

NeISF: Neural Incident Stokes Field for Geometry and Material Estimation, Chenhao Li, Taishi Ono, Takeshi Umemori, Hajime Mihara, Alexander Gatto, Hajime Nagahara, Yusuke Moriuchi

Non-Rigid Structure-from-Motion: Temporally-Smooth Procrustean Alignment and Spatially-Variant Deformation Modeling, Jiawei Shi, Hui Deng, Yuchao Dai

Small Steps and Level Sets: Fitting Neural Surface Models with Point Guidance, Chunmiao Hwna Konepugotadage, Yizhak Ben-Shabat, Dylan Campbell, Stephen Gould

CVT-xRF: Contrastive In-Voxel Transformer for 3D Consistent Radiance Fields from Sparse Inputs, Yiqiang Zhong, Lanqing Hong, Zhenguo Li, Dan Xu

GaussianEditor: Swift and Controllable 3D Editing with Gaussian Splatting, Yiwen Chen, Zilong Chen, Chi Zhang, Feng Wang, Xiaofeng Yang, Yikai Wang, Zhongang Cai, Lei Yang, Huaping Liu, Guosheng Lin

Cam4DOcc: Benchmark for Camera-Only 4D Occupancy Forecasting in Autonomous Driving Applications, Junyi Ma, Xieyuani Chen, Jiawei Huang, Jingyi Xu, Zhen Luo, Jintao Xu, Weihao Gu, Rui Ai, Hesheng Wang

UDIFF: Generating Conditional Unsigned Distance Fields with Optimal Wavelet Diffusion, Junsheng Zhou, Weiqi Zhang, Baorui Ma, Kano Shi, Yu-Shen Liu, Zhizhong Han

PanoRecon: Real-Time Panoptic 3D Reconstruction from Monocular Video, Dong Wu, Zika Yan, Hongsbin Zha

Three Pillars Improving Vision Foundation Model Distillation for Lidar, Gilles Puy, Spyros Gidaris, Alexandre Boulch, Oriane Siméoni, Corentin Saüter, Patrick Pérez, Andrei Bursuc, Renaud Marlet

GARField: Group Anything with Radiance Fields, Chung Min Kim, Mingxuan Wu, Justin Kerr, Ken Goldberg, Matthew Tancik, Angjoo Kanazawa

Flexible Depth Completion for Sparse and Varying Point Densities, Jinyung Park, Yu-Jhe Li, Kris Kitani

ReconFusion: 3D Reconstruction with Diffusion Priors, Rundi Wu, Ben Mildenhall, Philipp Henzler, Keunhong Park, Ruiqi Gao, Daniel Watson, Pratul P. Srinivasan, Dor Verbin, Jonathan T. Barron, Ben Poole, Aleksander Holynski

GLACE: Global Local Accelerated Cordinate Encoding, Fangjiyhua Wang, Xudong Jiang, Silvano Galliani, Christoph Vogel, Marc Pollefeys

NARUTO: Neural Active Reconstruction from Uncertain Targets, Observations, Ziyou Feng, Huanying Zhang, Zhen Chen, Qianan Yan, Xiangyu Xu, Changjiang Cai, Bing Li, Qilun Zhu, Yi Xu

Photo-SLAM: Real-time Simultaneous Localization and Photorealistic Mapping for Monocular Stereo and RGB-D Cameras, Huajian Huang, Longwei Li, Hui Cheng, Sai-Kit Yeung

Detector-Free Structure from Motion, Xingyi He, Jiaying Sun, Yifan Wang, Sida Peng, Qixing Huang, Hujun Bao, Xiaowei Zhou

Memory-based Adapt ids for Online 3D Scene Perception, Xiuxue Xu, Chong Xia, Zhiwei Wang, Qingfeng Zhao, Yaqion Duan, Jie Zhou, Jiwen Lu

SurroundSDF: Implicit 3D Scene Understanding Based on Signed Distance Field, Lizhe Liu, Bohua Wang, Hongwei Xie, Daqi Li, Li Liu, Zhiqiang Tian, Kuiyuan Yang, Bing Wang

CoGS: Controllable Gaussian Splatting, Heng Yu, Joel Julin, Zoltán A. Milacski, Koichiro Niinuma, László A. Jeni

DrivingGaussian: Composite Gaussian Splatting for Surrounding Dynamic Autonomous Driving Scenes, Xiaoyu Zhou, Zhiwei Liu, Xiaojun Shan, Yongtao Wang, Deqing Sun, Ming-Hsuan Yang

GS-IR: 3D Gaussian Splatting for Inverse Rendering, Zhihao Liang, Qi Zhang, Ying Feng, Ying Shan, Kui Jia

Cross-spectral Gated-RGB Stereo Depth Estimation, Samuel Brucker, Stefanie Waltz, Mario Bijelic, Felix Heide

Efficient LoFTR: Semi-Dense Local Feature Matching with Sparse-Like Speed, Yifan Wang, Xingyi He, Sida Peng, Dongli Tan, Xiaowei Zhou


Dynamic Cues-Assisted Transformer for Robust Point Cloud Registration, Hong Chen, Pei Yan, Sihe Xiang, Yihua Tan

Learning to Produce Semi-dense Correspondences for Visual Localization, Khang Truong Giang, Soohwan Song, Sungjo Ho

GP-NeRF: Generalized Perception NeRF for Context-Aware 3D Scene Understanding, Hao Li, Dingwen Chang, Yalin Dai, Nian Liu, Lechao Chen, Jingfeng Li, Jingdong Wang, Junwei Han

Compact 3D Gaussian Representation for Radius Field, Joo Chan Lee, Daniel Rho, Xiangyu Sun, Jong Hwan Ko, Eunyoung Park

Unsupervised Occupancy Learning from Sparse Point Cloud, Amine Ouasfi, Adnane Boukhayma

Grounding and Enhancing Grid-based Models for Neural Fields, Zelin Zhao, Fenglei Fan, Wenlong Liao, Junchi Yan

TACO: Benchmarking Generalizable Bimanual Tool-ACTION Object Understanding, Yun Lii, Haolin Yang, Xu Si, Ling Liu, Zipeng Li, Yuxiang Zhang, Yebin Liu, Li Yi

ImageNet-D: Benchmarking Neural Network Robustness on Diffusion Synthetic Object, Chenshuang Zhang, Fei Pan, Junno Kim, In So Kweon, Chengzhao Mao

SynFog: A Photo-realistic Synthetic Fog Dataset based on End-to-end Imaging Simulation for Advancing Real-World Defogging in Autonomous Driving, Yiming Xie, Henglu Wei, Zhenyi Liu, Xiaoyu Wang, Xiangyang Ji


Infinigen Indoors: Photorealistic Indoor Scenes using Procedural Generation, Alexander Raistrick, Lingjie Mei, Karhan Kayan, David Yan, Yiming Zuo, Binjing Han, Hongyu Wen, Meenal Parakh, Stamatis Alexandropoulos, Lahav Lipson, Zeyu Ma, Jia Deng

Probing the 3D Awareness of Visual Foundation Models, Mohamed El Banati, Amit Raj, Kevis-Kokitsu Maninis, Abhishek Kar, Yuanzhen Li, Michael Rubinstein, Deqing Sun, Leonidas Guibas, Justin Johnson, Varun Jampani

VDBench: Comprehensive Benchmark Suite for Video Generative Models, Ziqi Huang, Yinan He, Jiashuo He, Fan Zhang, Chenyang Si, Yiming Yuan, Zuo Binjing Han, Hongyu Wen, Meiinal Parakh, Stamatis Alexandropoulos, Lahav Lipson, Zeyu Ma, Jia Deng

MAPLM: A Real-World Large-Scale Vision-Language Benchmark for Map and Traffic Scene Understanding, Xu Cao, Tong Zhou, Yunseng Ma, Weni Yan, Can Cui, Kun Tang, Zhipeng Cao, Kaizhao Liang, Ziran Wang, James M. Rehg, Chao Zheng

Video Recognition in Portrait Mode, Mingfei Han, Linjie Yang, Xiaoqin Jin, Jiashi Feng, Xiaojun Chang, Heng Wang
222 MMVP: A Multimodal MoCap Dataset with Vision and Pressure Sensors, He Zhang, Shenghao Ren, Haolei Yuan, Jianhui Zhao, Fan Li, Shuangpeng Sun, Zhenghao Liang, Tao Yu, Qiu Shen, Xun Cao

223 What If the TV Was Off? Examining Counterfactual Reasoning Abilities of Multi-modal Language Models, Letian Zhang, Xiaotong Zhai, Zhongkai Zhao, Yongshuo Zong, Xin Wen, Bingchen Zhao

224 COCONut: Modernizing COCO Segmentation, Xueqing Deng, Qihang Yu, Peng Wang, Xiaohui Shen, Liang-Chieh Chen

225 Traffic Scene Parsing through the TSP6K Dataset, Peng-Tao Jiang, Yuqi Yang, Yang Cao, Qibin Hou, Ming-Ming Cheng, Chunhua Shen


227 Rethinking the Evaluation Protocol of Domain Generalization, Han Yu, Xinxuqin Zhang, Renzhe Xu, Jiahuo Liu, Yue He, Peng Cui

228 MMSum: A Dataset for Multimodal Summarization and Thumbnail Generation of Videos, Jieliin Qiu, Jiacheng Zhu, William Han, Aditesh Karthik Mittal, Claire Jin, Zhengyuan Yang, Linjie Li, Jianfeng Wang, Ding Zhao, Bo Li, Liujuan Wang

229 Learning from Synthetic Human Group Activities, Che-Jui Chang, Danrui Li, Deep Patel, Parth Goel, Honglu Zhou, Seonghyeon Moon, Samuel S. Sohn, Sejong Yoon, Vladimir Pavlovic, Mubbasir Kapadia

230 Instance Tracking in 3D Scenes from Egocentric Videos, Yunhan Zhao, Haoyu Ma, Shu Kong, Charless Fowlkes

231 Insect-Foundation: A Foundation Model and Large-scale 1M Dataset for Visual Insect Understanding, Hoang-Quan Nguyen, Thanh-Dat Truong, Xuan Bac Nguyen, Ashley Dowling, Xin Li, Khoa Luu

232 Low-Resource Vision Challenges for Foundation Models, Yunhua Zhang, Hazel Doughty, Cees G. M. Snoek


234 FreeMan: Towards Benchmarking 3D Human Pose Estimation under Real-World Conditions, Jiong Wang, Fengyu Yang, Bingliang Li, Wenbo Gou, Danqi Yan, Ailing Zeng, Yijun Li, Dunle Wang, Yangqing Jing, Ruiyuan Zhang

235 LiDAR-Net: A real-scanned 3D Point Cloud Dataset for Indoor Scenes, Yanwen Guo, Yunqian Ren, Dayong Ren, Xiaohong Zhang, Jiawei Li, Liang Pu, Hanbo Jin, Xuan Guo, Yiqun Wu, Yuyang Yang, Zhihao Ren, Xiaojun Wang, Hanyi Sun, Shuhui Cai, Zexing Xu, Mengnan Zhao


237 View-decoupled Transformer for Person Re-identification under Aerial-ground Camera Network, Quan Zhang, Lei Wang, Vishal M. Patel, Xiaohua Xie, Jianhui Lai, Tongtong Yuan, Xuange Xu, Jiyoon Kim, Hyung-Sin Kim

238 UFineBench: Towards Text-based Person Retrieval with Ultra-fine Granularity, Jialong Zuo, Hanyu Zhou, Ying Nie, Feng Zhang, Tianyu Guo, Nong Sang, Yunhai Wang, Changxin Gao

239 Towards Automatic Power Battery Detection: New Challenge and Benchmark, Tong Wang, Yunguang Li, Zhenyu Chen, Qian Yu, Lihe Zhang, Hanqi Liu, Jiaming Zuo, Yue Li, Jun Le, Yizhou Guo, Raymond Chan, Ying Shan

240 Localization Is All You Evaluate: Data Leakage in Online Mapping Datasets and How to Fix It, Adam Lilja, Junsheng Fu, Erik Stenborg, Lars Hammarstrand

241 DL3DV-10K: A Large-Scale Scene Dataset for Deep Learning-based 3D Vision, Lu Lin, Yichen Sheng, Zhi Tu, Wenzian Zhao, Cheng Xin, Kun Wan, Lantao Yu, Qianyu Guo, Zixun Yu, Yawen Lu, Yuanmao Li, Xingpeng Sun, Rohan Ashok, Aniruddha Mukherjee, Hao Kang, Xiangrui Kong, Gang Hua, Tianyi Zhang, Bedrich Benes, Aniket Bera

242 OmniMedVQA: A New Large-Scale Comprehensive Evaluation Benchmark for Medical VLMLM, Yu Tao, Tianbin Li, Quanfeng Lu, Wenqi Shao, Junjun He, Yu Qiao, Ping Luo

243 Can Biases in ImageNet Models Explain Generalization?, Paul Gavrylov, Janis Keuper

244 MVBench: A Comprehensive Multi-modal Video Understanding Benchmark, Kunchang Li, Yali Wang, Yinan He, Yizhuo Li, Yi Wang, Yi Liu, Jun Wang, Jian Xu, Guo Chen, Ping Luo, Limin Wang, Yu Qiao


246 Point-VOS: Pointing Up Video Object Segmentation, Sabarinath Mahadevan, Idil Esen Zulfikar, Paul Voigtlaender, Bastian Leibe

247 GPT-4(ison) is a Human-Aligned Evaluator for Text-to-3D Generation, Tong Wu, Guandao Yang, Zhibing Li, Kai Zhang, Ziwei Liu, Leonidas Guibas, Dahua Lin, Gordon Wetzstein

248 ConCon-Chi: Concept-Context Chimera Benchmark for Personalized Vision-Language Tasks, Andrea Rosasco, Stefano Berti, Giulia Pasquale, Damiano Malafonte, Shojo Sato, Hiroyuki Segawa, Tetsugou Inada, Lorenzo Natale

249 FISBe: A Real-World Benchmark Dataset for Instance Segmentation of Long-Range Thin Filamentous Structures, Lisa Mais, Peter Hirsch, Claire Managan, Ramya Kandarpa, Jose Lorenz Rumberger, Anikke Reineke, Lena Maier-Herin, Gudrun Ihrike, Dagmar Kainmueller

250 Inter-X: Towards Versatile Human-Human Interaction Analysis, Liang Xu, Xintao Lv, Yichao Yan, Xin Jin, Shuwen Wu, Congsheng Xu, Yifan Liu, Yizhou Zhou, Fengyun Rao, Xingdong Sheng, Yunhui Liu, Wenjun Zeng, Xiangkang Yang

251 TextNeRF: A Novel Scene-Text Image Synthesis Method based on Neural Radiance Fields, Jialei Cui, Jianwei Du, Wenzhuo Liu, Zhouhui Li

252 Systematic Comparison of Semi-supervised and Self-supervised Learning for Medical Image Classification, Zhe Huang, Ruijie Jiang, Shuchin Aeran, Michael C. Hughes

253 Unexplored Faces of Robustness and Out-of-Distribution: Covariate Shifts in Environment and Sensor Domains, Eunsu Baek, Keonho Park, Jiyoung Kim, Hyung-Sin Kim
PROGRAM GUIDE
FRIDAY, JUNE 21

Chao, Yu Su

299 A Noisy Elephant in the Room: Is Your Out-of-Distribution Detector Robust to Label Noise?, Gaëdricelle Humbert-Renaux, Sergio Escalera, Thomas B. Moeslund

300 eTraM: Event-based Traffic Monitoring Dataset, Aayush Atul Verma, Bhаратesh Chakravarti, Arpitsinh Vagheila, Hua Wei, Yezhou Yang


302 MSU-4S - The Michigan State University Four Seasons Dataset, Daniel Kent, Mohammed Alyaoubq, Xiaohu Lu, Hamed Khatoonabadi, Kookjin Sung, Cole Scheller, Alexander Dalat, Asma bin Thabit, Roberto Whitley, Hayder Radha

303 TUMTraF V2X Cooperative Perception Dataset, Walter Zimmer, Gerhard Arja Wardana, Suren Srinathran, Xincheng Zhou, Rui Song, Alois C. Knoll


305 Towards Co-Evaluation of Cameras HDR and Algorithms for Industrial-Grade 6DoF Pose Estimation, Agastya Kalra, Guy Stoppi, Dmitri Marin, Vage Taamazyan, Aarrushi Shandilya, Rishav Agarwal, Anton Boykov, Tze Hao Chong, Michael Stark

306 Scaling Laws for Data Filtering – Data Curation cannot be Compute Agnostic, Sachin Goyal, Pratyush Maini, Zachary C. Lipton, Aditi Raghunathan, J. Zico Kolter

307 Benchmarking Audio Visual Segmentation for Long-Untrimmed Videos, Chen Liu, Peike Patrick Li, Qingtao Yu, Hongwei Sheng, Dadong Wang, Lincheng Li, Xin Yu

308 MLP Can Be A Good Transformer Learner, Shao Lin, Pumeng Lyu, Congrui Liu, Tao Tang, Xiaodan Liang, Andy Song, Xiaojun Chang

309 From SAM to CAMs: Exploring Segment Anything Model for Weakly Supervised Semantic Segmentation, Hyeokjun Kweon, Jurin Yoon

310 Domain-Specific Block Selection and paired-view pseudo-labeling for online test-time adaptation, Yeonguk Yu, Sungho Shin, Seunghyeok Back, Nam Sung Kim, Kyoobin Lee

311 VideoMAC: Video Masked Autoencoders Meet Vision Models, GLID: Pre-training a Generalist Encoder-Decoder Vision Model, Hossam Isack, Abhishek Kar, Helge Rhodin, Andrea Tagliasacchi, Kwang Moo Yi

313 Rethinking the Representation in Federated Unsupervised Learning with Non-IID Data, Xinting Liao, Weiming Liu, Chaochao Chen, Pengyong Zhou, Fengyuan Yu, Huaibo Zhu, Binhui Yao, Tao Wang, Xiaolin Zheng, Yanchao Tan

312 GLID: Pre-training a Generalist Encoder-Decoder Vision Model, Jihao Liu, Jiacheng Zhang, Yu Liu, Hongsheng Li

313 Sequential Modeling Enables Scalable Learning for Large Vision Models, Yutong Bai, Xinyang Geng, Kartikkeya Mangalam, Amir Bar, Alan L. Yuille, Trevor Darrell, Jitendra Malik, Alexei A. Efros


316 CroSel: Cross Selection of Confident Pseudo Labels for Partial-Label Learning, Shiyu Tian, Hongxin Wei, Yiqun Wang, Lei Feng

317 BEM: Balanced and Diversity-based Mix for Long-Tailed Semi-Supervised Learning, Hongwei Zheng, Linyuan Zhou, Han Li, Jinming Su, Xiaoming Wei, Xiaoming Xu

318 ReCoRe: Regularized Contrastive Representation Learning of World Model, Rudra P. Koulou, Harit Pandya, Stephan Liwicki, Roberto Cipolla

319 Universal Novelty Detection Through Adaptive Contrastive Learning, Hossein Mirzaei, Mojtaba Nafez, Mohammad Jafari, Mohammad Bagher Soltani, Mohammad Azimzalayeri, Jafar Habibi, Mohammad Sabokrou, Mohammad Hossein Rohban

320 Learning to Count without Annotations, Lukas Knobel, Tengda Han, Yuki M. Asano

321 PointCloudPre-training with Diffusion Models, Xiao Zheng, Xiaoshui Huang, Guofeng Mei, Yuenan Hou, Zhaoyang Lyu, Bo Dai, Wanli Ouyang, Yongshun Gong

322 Improving Unsupervised Hierarchical Representation with Reinforcement Learning, Ruyi An, Yewen Li, Xu He, Pengjie Gu, Mengchen Zhao, Dong Li, Jianye Hao, Chaojie Wang, Bo An, Mingyuan Zhou

323 Investigating and Mitigating the Side Effects of Noisy Views for Self-Supervised Clustering Algorithms in Practical Multi-View Scenarios, Jie Xu, Yuzhou Ren, Xiaolong Wang, Lei Feng, Zheng Zhang, Gang Niu, Xiaofeng Zhu

324 Self-Supervised Representation Learning from Arbitrary Scenarios, Zhaowen Li, Yousong Zhu, Zhiyang Chen, Zongxin Gao, Rui Zhao, Chaoyang Zhao, Ming Tang, Jinqiao Wang

325 Learning SO(3)-Invariant Semantic Correspondence via Local Shape Transform, Chunhuyun Pan, Seungwook Kim, Jaesik Park, Minsu Cho

326 A Bayesian Approach to OOD Robustness in Image Classification, Prakhar Kaushik, Adam Kortylewski, Alain Yuille

327 Sculpting Holistic 3D Representation in Contrastive Language-Image 3D Pre-training, Yipeng Gao, Zeyu Wang, Wei-Shi Zheng, Cichang Xie, Yuyin Zhou

328 Solving Masked Jigsaw Puzzles with Diffusion Vision Transformers, Jinyang Liu, Wondmgezahu Teshome, Sandesh Ghimire, Mario Sznaier, Octavia Camps

329 DS-NeRV: Implicit Neural Video Representation with Decomposed Static and Dynamic Codes, Hao Yan, Zhihui Ke, Xiaobo Zhou, Tie Qiu, Xidong Shi, Dadong Jiang

330 Brain Decodes Deep Nets, Bin Shi, Jianfei Ruan, Tianze Pan, Hamed Alyaqoub, Xiaohu Lu, Xingzhe He, Hossam Isack, Abhishek Kar, Helge Rhodin, Andrea Tagliasacchi, Eric Hedlin, Gopal Sharma, Shweta Mahajan, Xingzhe He, Kwang Moo Yi

331 Distributionally Generative Augmentation for Fair Facial Attribute Classification, Fengda Zhang, Qianpei He, Kun Kuang, Jiashuo Liu, Long Chen, Chao Wu, Jun Xiao, Hanwang Zhang

332 Estimating Noisy Class Posterior with Part-level Labels for Noisy Label Learning, Rui Zhao, Bin Shi, Jianfei Ruan, Tianze Pan, Bo Dong

333 Improving Keypoints from Pretrained Diffusion Models, Eric Hedlin, Gopal Sharma, Shweta Mahajan, Xingzhe He, Hossam Isack, Abhishek Kar, Helge Rhodin, Andrea Tagliasacchi, Kwang Moo Yi


335 Mitigating Object Dependencies: Improving Point Cloud Self-Supervised Learning through Object Exchange, Yanhao Wu, Tong Zhang, Wei Ke, Congpei Qiu, Sabine Süsstrunk, Mathieu Salzmann
Adaptive Slot Attention: Object Discovery with Dynamic Slot Number, Ke Fan, Zechen Bai, Tianjun Xiao, Tong He, Max Horn, Yanwei Fu, Francesco Locatello, Zheng Zhang

Targeted Representation Alignment for Open-World Semi-Supervised Learning, Ruixuan Xiao, Lei Feng, Kai Tang, Junbo Zhao, Yixuan Li, Gang Chen, Haobo Wang

Hierarchical Correlation Clustering and Tree Preserving Embedding, Morteza Haghir Chehreghani, Mostafa Haghir Chehreghani

Contrastive Mean-Shift Learning for Generalized Category Discovery, Sui Choi, Dahyun Kang, Minsu Cho

CvULER: Enhanced Unsupervised Object Discoveries through Exhaustive Self-Supervised Transformers, Shahaf Arica, Or Rubin, Sapir Gershov, Shlomi Lauffer


HPL-ESS: Hybrid Pseudo-Labeling for Unsupervised Event-based Semantic Segmentation, Linglin Jing, Yiming Ding, Yunpeng Gao, Zhiqiang Wang, Xu Yan, Dong Wang, Gerald Schaefer, Hui Fang, Bin Zhao, Xuelong Li

Positive-Unlabeled Learning by Latent Group-Aware Meta Disambiguation, Lin Long, Haobo Wang, Zhijie Jiang, Lei Feng, Chang Yao, Gang Chen, Junbo Zhao

Aligning Logs Generatively for Principled Black-Box Knowledge Distillation, Jing Ma, Xiang Xian, Ke Wang, Yuchuan Wu, Yongbin Li

Improving Semantic Correspondence with Viewpoint-Guided Spherical Maps, Octave Mariotti, Oisin Mac Aodha, Hakan Bilen

Neural Modes: Self-supervised Learning of Nonlinear Modal Subspaces, Xiaohong Wang, Yinwei Du, Stelian Coro, Bernhard Thomaszewski

Decentralized Directed Collaboration for Personalized Federated Learning, Yingqi Liu, Yifan Shi, Qinglin Li, Baoyuan Wu, Xueqian Wang, Li Shen

Improving Graph Contrastive Learning via Adaptive Positive Sampling, Jingming Zhuo, Feiyang Qin, Can Cui, Kun Fu, Bingxin Niu, Mengzhu Wang, Yuanfang Guo, Chuan Wang, Zhen Wang, Xiaochun Cao, Liang Yang

Integrating Efficient Optimal Transport and Functional Maps For Unsupervised Shape Correspondence Learning, Tung Le, Khai Nguyen, Shaulin Sun, Nhat Ho, Xiaohui Xie

Unsupervised Feature Learning with Emergent Data-Driven Prototypicity, Yunhui Guo, Youren Zhang, Yubei Chen, Stella X. Yu

Label Propagation for Zero-shot Classification with Vision-Language Models, Vladan Stojnic, Yannis Kalantidis, Giorgos Tolias

Boosting Continual Learning of Vision-Language Models via Mixture-of-Experts Adapters, Jiaozu Yu, Yunzi Zhuge, Lu Zhang, Ping Hu, Dong Wang, Huchuan Lu, You He

Backpropagation-free Network for 3D Test-time Adaptation, Yanzhuo Wang, Ali Cheraghian, Zeeshan Hayder, Jie Hong, Sameera Ramasinghe, Shafin Rahman, David Ahmedtd-Aristizabal, Xuesong Li, Lars Petersson, Mehrdad Harandi

GDA: Generalized Diffusion for Robust Test-time Adaptation, Yun-Yun Tsai, Fu-Chen Chen, Albert Y. C. Chen, Junfeng Yang, Che-Chun Su, Min Sun, Cheng-Hao Kuo

Semantically-Shifted Incremental Adapter-Tuning is A Continual ViTransformer, Yuwen Tan, Qinbao Zhou, Xiang Xiang, Ke Wang, Yuchuan Wu, Yongbin Li

Few-shot Learner Parameterization by Diffusion Time-steps, Zhongyi Yue, Pan Zhou, Richang Hong, Hanwang Zhang, Qianru Sun

FREE: Faster and Better Data-Free Meta-Learning, Xiongxiang Wei, Zixuan Hu, Zhenyi Wang, Li Shen, Chun Yuan, Dacheng Tao

Classes Are Not Equal: An Empirical Study on Image Recognition Fairness, Jiequan Cui, Beiher Zhu, Xin Wen, Xiaojuan Qi, Bei Yu, Hanwang Zhang

DAVE - A Detect-and-Verify Paradigm for Low-Shot Counting, Jer Pelhan, Alan Lukežič, VitJan Zavrtanik, Matej Kristan


AMU-Tuning: Effective Logit Bias for CLIP-based Few-shot Learning, Yuwei Tang, Zhenyi Lin, Qiong Wang, Pengfei Zhu, Qinghua Hu

LEAD: Learning Decomposition for Source-free Universal Domain Adaptation, Sanying Gu, Tianpei Zou, Lianghua He, Florian Röhrtien, Alois Knoll, Guang Chen, Changqin Jiang

Improving Generalized Zero-Shot Learning by Exploring the Diverse Semantics from External Class Names, Yapeng Li, Yong Luo, Zengmao Wang, Bo Du

What When and When Should Object Detectors Update in Continually Changing Test Domains?, Jieyoon Yoo, Dongkwan Lee, Inseop Chung, Donghyun Kim, Nojun Kwak

Split to Merge: Unifying Separated Modalities for Unsupervised Domain Adaptation, Xinyao Li, Yuke Li, Zhekai Du, Fengling Li, Ke Lu, Jingjing Li

Domain-Agnostic Mutual Prompting for Unsupervised Domain Adaptation, Zhekai Du, Xinyao Li, Fengling Li, Ke Lu, Lei Zhu, Jingjing Li

Improving the Generalization of Segmentation Foundation Model under Distribution Shift via Weakly Supervised Adaptation, Haojie Zhang, Yongyi Su, Xun Xu, Kui Jia


Unified Language-driven Zero-shot Domain Adaptation, Senqiao Yang, Zhuotao Tian, Li Jiang, Jia Yia

Stable Neighbor Denoising for Source-free Domain Adaptive Segmentation, Dong Zhao, Shuang Wang, Qi Zang, Licheng Jiao, Nicu Sebe, Zhun Zhong

A Simple Recipe for Language-guided Domain Generalized Segmentation, Mohammad Fathes, Tuan-Hung Vu, Andrei Bursuc, Patrick Pérez, Raol de Charette

TCP:Textual-based Class-aware Prompt tuning for Visual-Language Model, Hantao Yao, Rui Zhang, Changsheng Xu

Adapters Strike Back, Jan-Martin O. Steitz, Stefan Roth

Improving Plasticity in Online Continual Learning via Collaborative Learning, Maorong Wang, Nicolas Michel, Ling Xiao, Toshihiko Yamasaki


Adaptive Random Feature Regularization on Fine-tuning Deep Neural Networks, Shiry Yamaguchi, Seikitoshi Kanai, Kazuki Adachi, Daiki Chijwa

ESCAPE: Encoding Super-keypoints for Category-Agnostic Pose Estimation, Xianpeng Yao, Xingchen Xian, Jiayen Li, Junjie Chen, Jiebin Yan, Kun Tian

PracticalDG: Perturbation Distillation on Vision-Language Models for Hybrid Domain Generalization, Zining Chen, Weiiqu Wang, Zhicheng Zhao, Fei Su, Aidong Men, Hongying Meng

Rethinking Multi-domain Generalization with A General Learning Objective, Zhaorui Tan, Xi Yang, Kaizhu Huang

L2B: Learning to Bootstrap Robust Models for Combating Label Noise, Yuyin Zhou, Xianhang Li, Fengze Liu, Yingyue Wei, Xuxi Chen, Lequan Yu, Chuang Xie, Matthew P. Langren, Lei Xing

Meta-Point Learning and Refining for Category-Agnostic Pose Estimation, Junjie Chen, Jiebin Yan, Yuming Fang, Li Niu

AZXP: Towards Private Domain Generalization, Geunhyeok Yu, Hooseok Hwang

Expandable Subspace Ensemble for Pre-Trained Model-Based Class-Incremental Learning, Da-Wei Zhou, Hai-Long Sun, Han-Jia Ye, De-Chuan Zhan

VRP-SAM: SAM with Visual Reference Prompt, Yanpeng Sun, Jiahui Chen, Shan Zhang, Xinyu Zhang, Qiang Chen, Gang Zhang, Errui Ding, Jingdong Wang, Zhichao Li

Flatten Long-Range Loss Landscapes for Cross-Domain Few-Shot Learning, Yixiong Zou, Yicong Liu, Yiman Hu, Yuhua Li, Ruixuan Li
3 ScribblePrompt: Fast and Flexible Interactive Segmentation for Any Biomedical Image, Hallee Wong, Marianne Rakic, John Guttag, Adrian Dalca
5 Visual Place Recognition using 3D City Models, Gabrielle Benton, Lorenz Junglas, Tom Polloco, Carlo Masone, Barbara Caputo
6 A Computer Vision Testbed for New York City Street Intersections, Mehmet Kerem Turkcan, Mahshid Ghasedehkordi, Sofia Kleisarchaki, Thomas Calkant, Levent Gürgen, Javad Ghaderi, Gil Zussman, Zoran Kostic
7 L-MAGIC: Language Model Assisted Generation of Images with Coherence, Zhipeng Cai, Tien Pei Chou
8 Building UBC in Minecraft, Ashtan Mistal
9 SuperPrimitive: Scene Reconstruction at a Primitive Level, Kirill Mazur, Guangbin Bae, Andrew J. Davison
10 H-Unique: 3D Hand Reconstruction and Automated Mapping of Anatomical Detail for Forensic Identification, Bryan M. Williams, Hossein Rahmani, Sue Black, Xinyu Yang, Zheheng Jiang, Andric Banica
11 Universal 3D Reconstruction: Interactive Demonstration of the Scalable 3D Lifting Foundation Model (3D-LFM), Mosam Dabhi, László A. Jeni, Simon Lucey
12 Neuro-Symbolic Olympics Diving Judge, Lauren Okamoto, Paritosh Parmar
13 Grounding Everything: Emerging Localization Properties in Vision-Language Transformers, Walid Bousselham
14 CoGS: Controllable Gaussian Splatting, Heng Yu, Joel Julin, Zoltan A Milascki, Koichiro Niinuma, László A. Jeni
15 Cutting-edge Text-Image Comprehension and Composition in Vision-Language Large Model, Jiaqi Wang, Xiaoyi Dong, Pan Zhang, Yuhang Zang
16 Collaborative Score Distillation for Consistent Visual Editing of My Own Visual Assets, Subin Kim, Sooyeon Park
17 Semantic Class-Adaptive Diffusion Model (SCA-DM), Alex Ergasti, Claudio Ferrari, Tomaso Fontanini, Massimo Bertozzi, Andrea Prati
18 A Real-Time Speech-Driven Vocal Tract Avatars, Tejas Prabhune, Peter Wu, Cheol Jun Cho, Bohan Yu, Gopala Anumanchipalli

12:00 - 14:00 LUNCH (Summit ExHall 1-2)

13:00 - 14:30 Orals 6A: Low-level Vision and Remote Sensing (Summit Ballroom)
1 LDP: Language-driven Dual-Pixel Image Defocus Deburring Network, Hao Yang, Liyuan Pan, Yan Yang, Richard Hartley, Miomiao Liu
2 S2MAE: A Spatial-Spectral Pretraining Foundation Model for Spectral Remote Sensing Data, Kuyang Li, Danteng Hong, Jocylyn Chanussot
3 Task-Driven Wavelets using Constrained Empirical Risk Minimization, Eric Marcus, Ray Sheombarsing, Jan-Jakob Sonke, Jonas Teuwen
4 Image Processing GNN: Breaking Rigidity in Super-Resolution, YuChuan Tian, Han Ting Chen, Chao Xu, Yunhe Wang

13:00 - 14:30 Orals 6B: Image & Video Synthesis (Summit Flex Hall AB)
1 Alchemist: Parametric Control of Material Properties with Diffusion Models, Prafull Sharma, Varun Jampani, Yuanzhen Li, Xuhui Jia, Dimitry Lagun, Fredo Durand, Bill Freeman, Mark Matthews
2 Generate Image Dynamics, Zhengqi Li, Richard Tucker, Noah Snively, Aleksander Holynski
3 Visual Anagrams: Generating Multi-View Optical Illusions with Diffusion Models, Daniel Geng, Inbum Park, Andrew Owens
4 MonoHair: High-Fidelity Hair Modeling from a Monocular Video, Keyu Wu, Lingchen Yang, Zhiyi Kuang, Yao Feng, Xutao Han, Yuefan Shen, Hongbo Fu, Kun Zhou, Youyi Zheng
5 Analyzing and Improving the Training Dynamics of Diffusion Models, Tero Karas, Milka Attalha, Jaakko Lehtinen, Janne Hellsten, Timo Ala, Samuli Laine

13:00 - 14:30 Orals 6C: Multi-Modal Learning (Summit Flex Hall C)
1 InternVL: Scaling up Vision Foundation Models and Aligning for Generic Visual-Linguistic Tasks, Zhe Chen, Jiannan Wu, Wenhai Wang, Weijie Su, Guo Chen, Sen Xing, Muyan Zhong, Qinglong Zhang, Xizhou Zhu, Lewei Lu, Bin Li, Ping Luo, Tong Yu, Qiao Ji, Jinfeng Dai
2 Describing Differences in Image Sets with Natural Language, Lina Dunlap, Yuhi Zhang, Xiaohan Wang, Ruiqi Zhang, Trevor Darrell, Jacob Steinhardt, Joseph E. Gonzalez, Serena Yeung-Levy
3 NoiseCLR: A Contrastive Learning Approach for Unsupervised Discovery of Interpretable Directions in Diffusion Models, Yusuf Dalva, Pinar Yanardag
4 MetaCloak: Preventing Unauthorized Subject-driven Text-to-image Diffusion-based Synthesis via Meta-learning, Yinlin Liu, Chenrui Fan, Yutong Dai, Xin Chen, Pan Zhou, Lichao Sun
5 EGTR: Extracting Graph from Transformer for Scene Graph Generation, Jinnbae Im, JeongYeoun Nam, Nokyung Park, Hyungmin Lee, Seunghyun Park

14:30 - 14:45 Courtesy Break
14:45 - 15:45 KEYNOTE 3 - Sofia Crespo, Artist (Summit Flex Hall ABC)
15:45 - 16:00 Courtesy Break
16:00 - 17:00 PANEL 2 (Summit Flex Hall ABC)
16:45 - 17:15 Poster Setup (Arch 4E)

17:15 - 18:00 Poster Session 6 & Exhibit Hall (Arch 4A-E)
1 MonoHair: High-Fidelity Hair Modeling from a Monocular Video, Keyu Wu, Lingchen Yang, Zhiyi Kuang, Yao Feng, Xutao Han, Yuefan Shen, Hongbo Fu, Kun Zhou, Youyi Zheng
2 BadCLIP: Trigger-Aware Prompt Learning for Backdoor Attacks on CLIP, Jiawang Bai, Kuofeng Gao, ShaoBo Min, Shu-Tao Xie, Zhifeng Li, Wei Liu
3 Semantic-Aware Multi-Label Adversarial Attacks, Hassan Mahmood, Ehsan Elhamifar
4 Defense without Forgetting: Continual Adversarial Defense with Anisotropic & Isotropic Pseudo Replay, Yuhang Zhou, Zhiyong Hu
5 Learning to Transform Dynamically for Better Adversarial Transferability, Rongyi Zhu, Zeliang Zhang, Susan Liang, Zhuo Liu, Chenliang Xu
6 Infrared Adversarial Car Stickers, Xaopei Zhu, Yuqi Liu, Zhanhao Hu, Jianmin Li, Xiaolin Hu
7 Unsegregate Anything by Simulating Deformation, Jiahao Lu, Xingyi Yang, Xinchao Wang
8 Efficient Model Stealing Defense with Noise Transition Matrix, Dong-Dong Wu, Chilin Fu, Weichang Wu, Wenwen Xia, Xiaolu Zhang, Jun Zhou, Min-Ling Zhang
9 Fully Exploiting Every Real Sample: SuperPixel Sample Gradient Model Stealing, Yunlong Zhao, Xiaoheng Dong, Yijing Liu, Xinjun Pei, Jiaxie Li, Wei Chen
10 Hide in Thicket: Generating Imperceptible and Rational Adversarial Perturbations on 3D Point Clouds, Tianrui Lou, Xiaojun Jia, Jindong Gu, Li Liu, Siyuan Liang, Bangyan He, Xiaochun Cao
11 Boosting Adversarial Transferability by Block Shuffle and Rotation, Kunyu Wang, Xuanran He, Wenxuan Wang, Xiaosong Wang
12 Robust Overfitting Does Matter: Test-Time Adversarial Purification with FGSM, Linyu Tang, Lei Zhang
13 Data Poisoning based Backdoor Attacks to Contrastive Learning, Jinghua Zhang, Hongbin Liu, Jinyuan Jia, Neil Zhenqiang Gong
14 NAPGuard: Towards Detecting Naturalistic Adversarial Patches, Siyang Wu, Jakai Wang, Jiejie Zhao, Yaxue Wang, Xianglong Liu
15 Ensemble Diversity Facilitates Adversarial Transferability, Bowen Tang, Zheng Wang, Yi Bin, Qi Dou, Yang Yang, Heng Tao Shen
16 Revamping Federated Learning Security from a Defender's Perspective: A Unified Defense with Homomorphic Encrypted Data Space, K. Naveen Kumar, Reshmi Mitra, C. Krishna Mohan
17 Can Protective Perturbation Safeguard Personal Data from Being Exploited by Stable Diffusion?, Zhengyue Zhao, Jinhao Duan, Kaidi Xu, Chenan Wang, Rui Zhang, Zidong Du, Qi Guo, Xing Hu
18 One Prompt Word is Enough to Boost Adversarial Robustness for Pre-trained Vision-Language Models, Lin Li, Haoyan Guan, Jianing Qiu, Michael Spratling
19 Watermark-embedded Adversarial Examples for Copyright Protection against Diffusion Models, Peifei Zhu, Tsubasa Takahashi, Hirokatsu Kataoka
20 Not All Prompts Are Secure: A Switchable Backdoor Attack Against Pre-trained Vision Transformers, Sheng Yang, Jiawang Bai, Kuofeng Gao, Yong Yang, Yiming Li, Shu-Tao Xia
21 Focus on Hiders: Exploring Hidden Threats for Enhancing Adversarial Training, Qian Li, Yuxiao Hu, Yinping Dong, Dongxiao Zhang, Yuntian Chen
22 Physical 3D Adversarial Attacks against Monocular Depth Estimation in Autonomous Driving, Junhao Zheng, Chenhao Lin, Jiahao Sun, Zhengyu Zhao, Qian Li, Chao Shen
23 Distraction is All You Need: Memory-Efficient Image Immunization against Diffusion-Based Image Editing Highlight: Top 10% of the accepted papers No Ling Lo, Cheng Yu Yeo, Hong-Han Shuai, Wen-Huang Cheng
24 PAD: Patch-Agnostic Defense against Adversarial Patch Attacks, Lihua Jing, Rui Wang, Wenqi Ren, Xin Dong, Cong Zou
25 PeerAID: Improving Adversarial Distillation from a Specialized Peer Tutor, Jaewon Jung, Hongsun Jang, Jaeyong Song, Jinhe Lee
26 Revisiting Adversarial Training Under Long-Tailed Distributions, Xing Xue, Ningping Mou, Qian Wang, Lingchen Zhao
27 Pre-trained Model Guided Fine-Tuning for Zero-Shot Adversarial Robustness, Sibo Wang, Jie Zhang, Zheng Yuan, Shiguang Shan
28 Towards Transferable Targeted 3D Adversarial Attack in the Physical World, Yao Huang, Yinpeng Dong, Shouwei Ruan, Xiao Yang, Hong Su, Xuexing Wei
29 Nearest is Not Dearest: Towards Practical Defense against Quantization-conditioned Backdoor Attacks, Boheng Li, Yishuo Cai, Haowei Li, Feng Xue, Zhifeng Li, Yiming Li
30 Perturbing Attention Gives You More Bang for the Buck: Subtle Imaging Perturbations That Efficiently Fool Customized Diffusion Models, Jingyao Xu, Yuetong Lu, Yandong Li, Siyuan Lu, Dongdong Wang, Xiang Wei
31 Boosting Adversarial Training via Fisher-Rao Norm-based Regularization, Xiangyu Yin, Wenjie Ruan
32 Random Entangled Tokens for Adversarially Robust Vision Transformer, Huilih Gong, Minjing Dong, Sqi Ma, Seyit Camtepe, Surya Nepal, Chang Xu
33 Backdoor Defense via Test-Time Detecting and Repairing, Jiyan Guan, Jian Liang, Ran He
34 1-Lipschitz Layers Compared: Memory Speed and Certifiable Robustness, Bernd Prach, Fabio Brau, Giorgio Buttazzo, Christoph H. Lampert
35 DiffAM: Diffusion-based Adversarial Makeup Transfer for Facial Privacy Protection, Yuhao Sun, Lingyun Yu, Hongtao Xie, Jiaming Li, Yongdong Zhang
36 DAP: A Dynamic Adversarial Patch for Evading Person Detectors, Amira Guesmi, Ruitian Ding, Muhammad Abdullah Hannif, Ihsen Alouani, Muhammad Shafique
37 Adversarial Distillation Based on Slack Matching and Attribution Region Alignment, Shenglin Yin, Zhen Xiao, Mingxuan Song, Jieyi Long
38 Improving Transferable Targeted Adversarial Attacks with Model Self-Enhancement, Han Wu, Guanyan Ou, Weibin Wu, Zibin Zheng
39 On the Robustness of Large Multimodal Models Against Image Adversarial Attacks, Xuanming Cui, Alejandro Aparcedo, Young Kyun Jung, Ser-Nam Lim
41 BadCLIP: Dual-Embedding Guided Backdoor Attack on iMultimodal Contrastive Learning, yuan Liang, Mingyi Zhu, Aishan Liu, Baoyuan Wu, Xiaochun Cao, En-Chien Chang
42 MMCert: Provably Defense against Adversarial Attacks to Multi-modal Models, Yanming Wang, Hongyu Fu, Wei Zou, Jinjun Jia
43 MimicDiffusion: Purifying Adversarial Perturbation via Mimicking Clean Diffusion Model, Kiyu Song, Hanjiang Lai, Yan Pan, Jian Yin
44 Revisiting Adversarial Training at Scale, Zeyu Wang, Xianhang Li, Hongru Zhu, Chiang Xie
45 Language-Driven Anchors for Zero-Shot Adversarial Robustness, Xiao Li, Wei Zhang, Yining Liu, Zhanhao Hu, Bo Zhang, Xiaolin Hu
46 Transferable Structural Sparse Adversarial Attack Via Exact Group Sparsity Training, Di Ming, Peng Ren, Yunlong Wang, Xin Feng
47 Fooling Polarization-Based Vision using Locally Controllable Polarization Projecting, Zhuoxiao Li, Zhihang Zhong, Shohei Nobuhara, Ko Nishino, Yinqiang Zheng
49 Attacking To Defend: Exploiting Adversarial Attacks for Detecting Poisoned Models, Samar Fares, Karthik Nandakumar
50 Towards Understanding and Improving Adversarial Robustness of Vision Transformers, Samyak Jain, Tanima Dutta
51 Towards Fairness-Aware Adversarial Learning, Yanhao Zhang, Tianle Zhang, Ronghui Ma, Xiaowei Huang, Wenjie Ruan
52 Byzantine-robust Decentralized Federated Learning via Dual-domain Clustering and Trust Bootstrapping, Peng Sun, Xinyang Liu, Zhibo Wang, Bo Liu
53 Towards General Robustness Verification of MaxPool-based Convolutional Neural Networks via Tightening Linear Approximation, Yuan Xiao, Shiqing Ma, Juan Zhai, Chunrong Fang, Jinjun Jia, Zhenyu Chen
54 Soften To Defend: Towards Adversarial Robustness via Self-Labeling Label Refinement, Zhourong Li, Daiwei Yu, Lina Wei, Canghong Jin, Yun Zhang, Sixian Chen
56 LOTUS: Evasive and Resilient Backdoor Attacks through Sub-Partitioning, Siyuan Cheng, Guanhong Tao, Yingyi Liu, Guanguy Shen, Shengwei An, Shiwei Feng, Xingzhe Xu, Kaiyuan Zhang, Shiqing Ma, Xinyu Zhang
57 Deep-TROI: An Inference Stage Trojan Insertion Algorithm through Efficient Weight Replacement Attack, Sabbir Ahmed, Ranyang Zhou, Shaahin Angizi, Adnan Siraj Rakin
59 Initialization Matters for Adversarial Transfer Learning, Andong Hua, Jindong Gu, Zhiyu Xue, Nicholas Carlini, Eric Wong, Yao Qin
60 Strong Transferable Adversarial Attacks via Ensembled Asymptotically Normal Distribution Learning, Zhengwei Fang, Rui Wang, Tao Huang, Liping Jing
61 HDRFlow: Real-Time HDR Video Reconstruction with Large Motions, Gangwei Xu, Yujin Wang, Jinwei Gu, Tianfan Xue, Xin Yang
62 A Physics-informed Low-rank Deep Neural Network for Blind and Universal Lens Aberration Correction, Jin Gong, Runzhao Yang, Weihang Zhang, Jinli Suo, Qionghai Dai
63 Super-Resolution Reconstruction from Bayer-Pattern Spike Streams, Yanchen Dong, Ruiqin Xiong, Jian Zhang, Zhaofei Yu, Xiaofeng Fan, Shuyuan Zhu, Tiejun Huang
64 In2SET: Intra-Inter Similarity Exploiting Transformer for Dual-Camera Compressive Hyperspectral Imaging, Xin Wang, Lizhi Wang, Xianglan Ma, Maqing Zhang, Lin Zhu, Hua Huang
66 Language-driven All-in-one Adverse Weather Removal, Hao Yang, Liuyuan Pan, Yan Yang, Wei Liang
Discovery of Interpretable Directions in Diffusion Models, Yusuf Dalva, Pinar Yanardag
Analyzing and Improving the Training Dynamics of Diffusion Models, Tero Karras, Miaa Aittala, Jaakko Lehtinen, Janne Hellsten, Timo Aila, Samuli Laine
Doubt Prior-Guided Diffusion for Zero-Shot Joint Low-Light Enhancement and Deblurring, Xiaoqian Lv, Shengping Zhang, Chenyang Wang, Yichen Zheng, Bineng Zhong, Chongyi Li, Liqiang Li
Color Shift Estimation-and-Correction for Image Enhancement, Yiyu Li, Ke Xu, Gerhard Peters, Hancke, Rynson W.H. Lau
Video Super-Resolution Transformer with Masked Inter-frame Attention, Xingyu Zhou, Leheng Zhang, Xiaorui Zhao, Keze Wang, Leida Li, Shuhang Gu
Distilling Semantic Priors from SAGM to Efficient Image Restoration Models, Quan Zhang, Xiaoyu Liu, Wei Li, Hanning Chen, Junchao Liu, Jie Hu, Zhwei Xiong, Chun Yuan, Yunhe Wang
Beyond Average: Individualized Visual Scanpath Prediction, Xianyu Chen, Ming Jiang, Qi Zhao
Multimodal Prompt Perceiver: Empower Adaptiveness Generalizability and Fidelity for All-in-One Image Restoration, Yang Ai, Huaibo Huang, Xiaoxiang Zhou, Jiexiang Wang, Ran He
Selective Hourglass Mapping for Universal Image Restoration Based on Diffusion Model, Dian Zheng, Xiao-Ming Wu, Shuzhou Yang, Jia Zhang, Jian-Fang Hu, Wei-Shi Zheng
SeeSR: Towards Semantics-Aware Real-World Image Super-Resolution, Rongyuan Wu, Tae Yoon, Lingchen Sun, Zhengqiang Zhang, Shuai Li, Lei Zhang
Revisiting Single Image Reflection Removal In the Wild, Yuri Zhu, Xueyang Fu, Peng-Tao Jiang, Hao Zhang, Qinbin Sun, Jinwei Chen, Zheng-Jun Zha, Bo Li
ODCR: Orthogonal Decoupling Contrastive Regularization for Unpaired Image Dehazing, Zhongze Wang, Haitao Zhao, Jingchao Peng, Lujian Yao, Kaijie Zhao
Q-Instruct: Improving Low-level Visual Abilities for Multi-modality Foundation Models, 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
Enhancing Quality of Compressed Images by Mitigating Enhancement Bias Towards Compression Domain, Quanliang Xing, Mai Xu, Shengxi Li, Xin Deng, Meisong Zheng, Huaida Liu, Ying Chen
Attentive Illumination Decomposition Model for Multi-Illuminant White Balancing, Dongyoung Kim, Jinwoo Kim, Junsang Yu, Seon Joo Kim
NightCC: Nighttime Color Constancy via Adaptive Channel Masking, Shuwei Li, Robby T. Tan
Navigating Beyond Dropout: An Intriguing Solution towards Generalizable Image Super Resolution, Hongjun Wang, Jiayuan Chen, Yinqiang Zheng, Tieyong Zeng
Learning Inclusion Matching for Animation Paint Bucket Colorization, Yuekun Dai, Shangchen Zhou, Qinyue Li, Chongyi Li, Chen Change Loy
Defense Against Adversarial Attacks on No-Reference Image Quality Models with Gradient Norm Regularization, Yujia Liu, Chenxi Yang, Dingquan Li, Jianhao Ding, Tingting Jiang
Towards Backward-Compatible Continual Learning of Image Compression, Zhihao Duan, Ming Lu, Justin Yang, Jiapeng He, Zhan Ma, Fengqing Zhu
APISR: Anime Production Inspired Real-World Anime Super-Resolution, Boyang Wang, Fengyu Yang, Xihang Yu, Chao Zhang, Hanbin Zhao
Unifying Automatic and Interactive Matting with Pretrained ViTs, Zixuan Ye, Wenzhe Liu, He Guo, Yujia Liang, Chaoyi Hong, Hao Lu, Zhiqiu Cao
Motion-adaptive Separable Collaborative Filters for Blind Motion Deblurring, Chengxi Liu, Xuan Wang, Xiangyu Xu, Ruhao Tian, Shuai Li, Xuejing Qian, Ming-Hsuan Yang
Genuine Knowledge from Practice: Diffusion-Test-Time Adaptation for Video Adverse Weather Removal, Yijun Yang, Hongtao Wu, Angelica L. Aviles-Rivero, Yukun Zhang, Jing Qin, Lei Zhu
HomoFormer: Homogenized Transformer for Image Shadow Removal, Jie Xiao, Xueyong Fu, Yurui Zhu, Dong Li, Jie Huang, Kai Zhu, Zheng-Jun Zha
Bidirectional Multi-Scale Implicit Neural Representations for Image Deraining, Xiang Chen, Jinhshan Pan, Jiangxin Dong
LED: A Large-scale Real-world Paired Dataset for Event Camera Denoising, Yuxing Duan
Seeing Motion at Nighttime with an Event Camera, Haoyue Liu, Shihan Peng, Lin Zhu, Yi Chang, Hanyu Zhou, Luxin Yan
Leveraging Frame Affinity for sRGB-to-RAW Video De-rendering, Chen Zhang, Wenchang Han, Yang Zhou, Jianbing Shen, Cheng-xu Zhu, Wentao Liu
Scaling Up to Excellence: Practicing Model Scaling for Photo-Realistic Image Restoration In the Wild, Fanghua Yu, Jinjing Gu, Zheyuan Li, Jinfan Hu, Xiangtiao Kong, Xintao Wang, Jingwen He, Yu Qiao, Chao Dong
AdaRevD: Adaptive Patch Exiting Reversible Decoder Pushes the Limit of Image Deblurring, Xintian Mao, Qingsi Li, Yan Wang
Unsupervised Blind Image Deblurring Based on Self-Enhancement, Lufei Chen, Xiangpeng Tian, Shuhua Xiong, Yinjie Lei, Chao Ren
TTA-EVF: Test-Time Adaptation for Event-based Video Frame Interpolation via Reliable Pixel and Sample Estimation, Hoonheo Cho, Taewoo Kim, Yuhwan Jeong, Kuk-Jin Yoon
Learning Coupled Dictionaries from Unpaired Data for Image Super-Resolution, Longguang Wang, Juncheng Li, Yingqiang Wang, Qingyong Hu, Yulan Guo
Empowering Resampling Operation for Ultra-High-Definition Image Enhancement with Model-Asserved Guidance, Wei Yu, Jie Huang, Bing Li, Kaiwen Zheng, Qi Zhu, Man Zhou, Feng Zhao
Generating Content for HDR Dehosting from Frequency View, Tao Hu, Qingsen Yan, Yuankai Qi, Yanning Zhang
Dual Prior Unfolding for Snapshot Compressive Imaging, Jiancheng Zhang, Hainin Zeng, Jiezhang Cao, Yongyong Chen, Dengxiu Yu, Yin-Ping Zhao
Binarized Low-light Raw Video Enhancement, Gengchen Zhang, Yulun Zhang, Xin Yuan, Ying Fu
Neural Spline Fields for Burst Image Fusion and Layer Separation, Ilya Chugunov, David Shustin, Ruyu Yan, Chenyang Lei, Felix Heide
Learning Degradation-Independent Representations for Camera ISP Pipelines, Yanhui Guo, Fangzhou Luo, Xiaolin Wu
SeD: Semantic-Aware Discriminator for Image Super-Resolution, Bingchen Li, Xin Li, Hanxin Zhu, Yeying Jin, Ruoyu Feng, Zhizheng Zhang, Zhibo Chen
SinSR: Diffusion-Based Image Super-Resolution in a Single Step, Yufei Wang, Wenhan Yang, Xinyuan Chen, Yajun Wang, Laping Guo, Lap-Pui Chau, Ziwei Liu, Yu Qiao, Alex C. Kot, Bihan Wen
Self-Adaptive Reality-Guided Diffusion for Artifact-Free Super-Resolution, Qingping Zheng, Ling Zheng, Yuanfan Guo, Ying Li, Songcen Xu, Jiankang Deng, Hang Xu
Improving Spectral Snapshot Reconstruction with Spectral-Spatial Rectification, Jiancheng Zhang, Hainin Zeng, Yongyong Chen, Dengxiu Yu, Yin-Ping Zhao
Diffusion-based Blind Text Image Super-Resolution, Yuzhe Zhang, Jiawei Zhang, Hao Li, Zhihui Wang, Luweii Hou, Dongqing Zou, Liheong Bian
CAMixerSR: Only Details Need More “Attention”, Yan Wang, Yi Li, Shijie Zhao, Junlin Li, Li Zhang
Diffusion-based BLurring AllGmentation, Jia-Hao Wu, Fu-Jen Tsai, Yan-Tsong Peng, Chunchi Tsai, Chia-Wen Lin, Yen-Yu Lin
Low-Res Leads the Way: Improving Generalization for Super-Resolution by Self-Supervised Learning, Haoyu Chen, Wenbo Li, Jinjing Gu, Jingjing Ren, Haoze Sun, Xueyi Zou, Zhenzong Zhang, Youliang Yan, Lei Zhu
CoSeR: Bridging Image and Language for Cognitive Super-Resolution, Haoze Sun, Wenbo Li, Jianhuang Liu, Haoyu Chen, Renjing Pei, Xueyi Zou, Youliang Yang, Yujia Yang
PROGRAM GUIDE


219 MultiPLY: A Multisensory Object-Centric Embodied Large Language Model in 3D World, Yining Hong, Zhihong Zheng, Peihao Chen, Yian Wang, Junyan Li, Chuang Gan

220 GPT4Point: A Unified Framework for Point-Language

221 Understanding and Generation, Zhangyang Qi, Ye Fang, Zezi Sun, Xiaoyang Wu, Tong Wu, Jiawei Wang, Dahua Lin, Hengyuan Zhu, Jayuan Fan, Tao Chen

222 Unified-IO 2: Scaling Autoregressive Multimodal Models with

223 Vision Language Audio and Action, Jiasen LU, Christopher Clark, Sangho Lee, Zichen Zhang, Savya Khosla, Ryan Marten, Derek Hoiem, Aniruddha Kembhavi

224 SHAP-EDITOR: Instruction-Guided Latent 3D Editing in Seconds, Qinghao Chen, Junyu Xie, Ira Laina, Andrea Vedaldi

225 Learning to Visually Localize Sound Sources from Mixtures without Prior Source Knowledge, Dongqin Kim, Sung Jin Im, Sangmin Jeon, Jung Uk Kim

226 Bring Event into RGB and LiDAR: Hierarchical Visual-Motion Fusion for Scene Flow, Hanyu Zhou, Yi Chang, Zhiwei Shi

227 Dispel Darkness for Better Fusion: A Controllable Visual Enhancer based on Cross-modal Conditional Adversarial Learning, Hao Zhang, Linfeng Tang, Xinyu Xiang, Xuhui Zuo, Jiayi Ma

228 Unraveling Instance Associations: A Closer Look for Audio-Visual Segmentation, Yuanhong Chen, Yuyuan Liu, Hu Wang, Fengbei Liu, Chong Wang, Helen Frazer, Gustavo Carneiro

229 DMR: Decomposed Multi-Modality Representations for Frames and Events Fusion in Visual Reinforcement Learning, Haoran Xu, Peizhi Peng, Guang Tan, Yuan Li, Xinhai Xu, Yonghong Tian

230 Text-Guided Variational Image Generation for Industrial Anomaly Detection and Segmentation, Mingyu Lee, Jongwon Choi

231 Tactile-Augmented Radiation Fields, Yiming Dou, Fengyu Yang, Yi Liu, Antonio Loquercio, Andrew Owens

232 LION: Empowering Multimodal Large Language Model with Dual-Level Visual Knowledge, Gongwei Chen, Leyang Shen, Rui Shao, Xiang Dang, Liqiang Nie

233 SDSTrack: Self-Distillation Symmetric Adapter Learning for Multi-Modal Visual Object Tracking, Xiaojun Hou, Jiazheng Xing, Yijie Qian, Yaowei Guo, Shuo Xin, Junhao Chen, Kai Tang, Mengmeng Wang, Zhengkai Jiang, Liang Liu, Yong Liu

234 Exploring the Transferability of Visual Prompting for Multimodal

235 Large Language Models, Yichun Zhang, Yinpeng Dong, Siyuan Zhang, Tianzhan Min, Hang Su, Jun Zhu

236 Maak Grounding for Referring Image Segmentation, Yang Xien Chng, Henry Zheng, Yizeng Han, Kuchong Qiu, Gao Huang

237 OneLLM: One Framework to Align All Modalities with Language, Jiaming Han, Kaixiiong Gong, Yiyuan Zhang, Jiaqi Wang, Kaipeng Zhang, Dahua Lin, Yu Qiao, Peng Gao, Xiangyu Yue


239 ModaVerse: Efficiently Transforming Modalities with LLMs, Xinyu Wang, Bohan Zhuang, Qi Wu

240 PromptKLD: Unsupervised Prompt Distillation for Vision-Language Models, Zheng Li, Xiang Li, Xinyi Fu, Xin Zhang, Weiqiang Wang, Shuo Chen, Jian Yang

241 Dynamic Prompt Optimizing for Text-to-Image Generation, Wenyi Mo, Tianyu Zhang, Yalong Bai, Bing Su, Ji-Long Wen, Qing Yang

242 Domain Prompt Learning with Quaternion Networks,

243 Qiong Cao, Zhengqin Xu, Yuntian Chen, Chao Ma, Xiaokang Yang

244 VIT-Lens: Towards Omni-modal Representations, Weixian Lei, Yixiao Ge, Kun Yi, Jianfeng Zhang, Difei Gao, Dylan Sun, Yuying Ge, Ying Shang, Mike Zheng Shou

245 Rotated Multi-Scale Interaction Network for Referring Remote Sensing Image Segmentation, Sihan Liu, Yiwei Ma, Xiaojing Zhang, Haowei Wang, Jiayi Ji, Xiaoshuai Sun, Rongrong Ji

246 Cyclic Learning for Binaural Audio Generation and Localization, Zhaoyu Li, Bin Zhao, Yuan Yuan

247 Learning to Rematch Mismatched Pairs for Robust Cross-Modal Retrieval, Haochen Han, Qinghua Zheng, Guan Dai, Minnan Luo, Jingdong Wang

248 VILA: On Pre-training for Visual Language Models, Ji Lin, Hongyu Yin, Wei Ping, Pavlo Molchanov, Mohammad Shoeybi, Song Han

249 A Picture is Worth More Than 77 Text Tokens: Evaluating CLIP-Style Models on Dense Captions, Jack Urbanek, Florian Bordes, Pietro Astolfi, Mary Williamson, Vasu Sharma, Adriana Romero-Soriano

250 How to Configure Good In-Context Sequence for Visual Question Answering, Li Li, Jiawei Peng, Huiyin Chen, Chongyang Gao, Xu Yang

251 CrossMAE: Cross-Modality Masked Autoencoders for Region-Aware Audio-Vision Pre-Training, Yuxin Guo, Siyang Sun, Shuailei Ma, Kecheng Zheng, Xiaoyi Bao, Shijie Ma, Wei Zou, Yun Zheng

252 Modality-Collaborative Test-Time Adaptation for Action Recognition, Baochen Xiong, Xiaoshan Yang, Yagoung Song, Yaowei Wang, Changsheng Xu

253 T-VEL: Text-Guided Visual Sound Source Localization in Mixtures, Tanvir Mahmud, Yapeng Tian, Diana Marinescu


255 Monkey: Image Resolution and Text Label Are Important Things

256 for Large multi-modal Models, Zhang Li, Biao Yang, Qiang Liu, Zhiyin Ma, Shuo Zhang, Jingyu Yang, Yabo Sun, Yuliang Liu, Xiang Bai

257 Rethinking Multi-view Representation Learning via Distilled Disentangling, Guanzhou Ke, Bo Wang, Xiaoli Wang, Shengfeng He

258 Causal Mode Multiplexer: A Novel Framework for Unbiased Multispectral Pedestrian Detection, Taehoon Kim, Sebin Shin, Youngsung Yu, Hak Gu Kim, Yong Man Ra


261 Efficient Vision-Language Pre-training by Cluster Masking, Zihao Wei, Zixuan Pan, Andrew Owens

262 MeLFusion: Synthesizing Music from Image and Language Cues

263 using Diffusion Models, anjoy Chowdhury, Sayan Nag, K J Joseph, Balalay Vasan Srinivasan, Dinesh Manocha

264 Weakly Misalignment-free Adaptive Feature Alignment for UAVs-based Multimodal Object Detection, Chen Chen, Jihaoo Qi, Xingyue Liu, Kangcheng Bin, Ruigang Fu, Xikun Hu, Ping Zhong

265 DIVAS: Video and Audio Synchronization with Dynamic Frame Rates, Clara Fernandez-Labrador, Mercan Ackar, Eitan Abecassis, Joan Massich, Christopher Schroers

266 Querying as Prompt: Parameter-Efficient Learning for Multimodal Language Model, Tian Liang, Jing Huang, Ming Kong, Luyuan Chen, Qiang Zhu

267 SonicVisionLM: Playing Sound with Vision Language Models, Zhifeng Xie, Shengye Yu, Qile He, Mengtian Li

268 Embracing Unimodal Aleatoric Uncertainty for Robust Multimodal Fusion, Zixian Gao, Xun Jiang, Xing Xu, Fumin Shen, Yujie Li, Heng Tao Shen

269 C3Net: Compound Conditioned ControlledNet for Multimodal Content Generation, Juntao Zhang, Yuehui Liu, Yu-Wing Tai, Chi-Keung Tang

270 Composed Video Retrieval via Enriched Context and Discriminative Embeddings, Omkar Thawar, Muzammal Naseer, Rao Muhammad Anwer, Salman Khan, Michael Felsberg, Mubarak Shah, Fahad Shakhbaz Khan

271 Looking Similar Sounding Different: Leveraging Counterfactual Cross-Modal Pairs for Audiovisual Representation Learning, Nikhil Singh, Chih-Wei Wu, Irooro Orife, Mahdi Kalaye

272 Anchor-based Robust Finetuning of Vision-Language Models, Jinwei Han, Zhiwen Lin, Zhongysun Lin, Yingguo Gao, Ke Yan, Shouhong Ding, Yuan Gao, Gui-Song Xia

FRIDAY, JUNE 21

63
268 Event-based Visible and Infrared Fusion via Multi-task Collaboration, Mingyueng Meng, Lin Zhu, Lizi Wang, Wei Zhang, Ruixing Xiong, Yonghong Tian
269 Prompt Learning via Meta-Regularization, Jinyoung Park, Juyeon Ko, Hyunwoo J. Kim
270 Knowledge-Enhanced Dual-stream Zero-shot Composed Image Retrieval, Yucheng Suo, Fan Ma, Linchao Zhu, Yi Yang
271 Contextual Augmented Global Contrast for Multimodal Intent Recognition, Kaili Sun, Zhiwen Xie, Meng Ye, Huyin Zhang
272 MRFS: Mutually Reinforcing Image Fusion and Segmentation, Hao Zhang, Xuhui Zuo, Jie Jiang, Chunshuo Guo, Jiayi Ma
273 POPDG: Popular 3D Dance Generation with PopDanceSet, Zhenyu Luo, Min Ren, Xuecai Hu, Yongzhen Huang, Li Yao
274 How to Make Cross Encoder a Good Teacher for Efficient Image-Text Retrieval?, Yuxin Chen, Zongyang Ma, Ziqi Zhang, Zhongang Qi, Chunfeng Yuan, Bing Li, Junfu Pu, Ying Shan, Xiaoqian Qi, Weiming Hu
275 Active Prompt Learning in Vision Language Models, Jiuhwan Bang, Sumyeong Ahn, Jae-Gil Lee
276 Descriptor and Word Soups: Overcoming the Parameter Efficiency Accuracy Tradeoff for Out-of-Distribution Few-shot Learning, Christopher Liao, Theodoros Tsiligkaridis, Brian Kulis
277 Text-IP: Leveraging Semantic Text Guidance for Degradation-Aware and Interactive Image Fusion, Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, Jiayi Ma
278 Hallucination Augmented Contrastive Learning for Multimodal Large Language Model, Chaoya Jiang, Haiyang Xu, Mengfan Dong, Jiaxing Chen, Wei Ye, Ming Yan, Qinghao Ye, Ji Zhang, Fei Huang, Shikun Zhang
279 Beyond Text: Frozen Large Language Models in Visual Signal Comprehension, Lei Zhu, Fangyun Wei, Yanyu Lu
280 Learning Spatial Features from Audio-Visual Correspondence in Egocentric Videos, Sagnik Majumder, Ziad Al-Halah, Karren Grauman
281 ES*: Evolving Self-Supervised Learning of Robust Audio-Visual Speech Representations, Yuanchang Zhang, Shuang Yang, Shiqiang Shan, Xinlin Chen
283 ULIP-2: Towards Scalable Multimodal Pre-training for 3D Understanding, Le Xue, Ning Yu, Shu Zhang, Artemis Panagopoulos, Junnan Li, Roberto Martin-Martín, Jiajiong Wu, Caiming Xiong, Ran Xu, Juan Carlos Niebles, Silvio Savarese
285 Language-aware Visual Semantic Distillation for Video Question Answering, Bo Zou, Chao Yang, Yu Qiao, Chengbin Quan, Youjian Zhao
286 PerceptionGPT: Effectively Fusing Visual Perception into LLM, Renjie Pi, Lewei Yao, Jiahui Gao, Jipeng Zhang, Tong Zhang
287 Cooperation Does Matter: Exploring Multi-Order Bilateral Relations for Audio-Visual Segmentation, Qi Yang, Xing Nie, Tong Li, Pengfei Gao, Ying Guo, Ching Chen, Pengfei Yan, Shiqiang Shan, Xinlin Chen
288 MV-Adapter: Multimodal Video Transfer Learning for Video Text Retrieval, Xiaojie Jin, Bowen Zhang, Weibo Gong, Kai Xu, Xueqin Deng, Peng Wang, Zhao Zhang, Xiaohui Shen, Jiashui Feng
289 Draw Step by Step: Reconstructing CAD Construction Sequences from Point Clouds via Multimodal Diffusion, Weijian Ma, Shuaiqi Chen, Yunzhong Lou, Yueyang Li, Xiandong Zhou
290 AV-RIR: Audio-Visual Room Impulse Response Estimation, Anton Ratnarajah, Sreyan Ghosh, Sonal Kumar, Purva Chiniya, Dinesh Manocha
291 Link-Context Learning for Multimodal LLLMs, Yan Tai, Weichen Fan, Zhao Zhang, Ziwei Liu
292 Unveiling the Power of Audio-Visual Early Fusion Transformers with Dense Interactions through Masked Modeling, Shenlong Mo, Pedro Morgado
293 Noisy-Correspondence Learning for Text-to-Image Person Re-identification, Yang Qin, Yingke Chen, Dezong Peng, Xi Peng, Joey Tianyi Zhou, Peng Hu
294 Mind Artist: Creating Artistic Snapshots with Human Thought, Jiaxuan Chen, Yu Qi, Yueming Wang, Gang Pan
297 Data-Efficient Multimodal Fusion on a Single GPU, Noël Voultis, Zhaoyan Liu, Satya Krishna Gorti, Valentijn Vluchroze, Jesse C. Cresswell, Guangwei Yu, Gabriel Loaiza-Ganem, Maksims Volkovs
298 SoundgActions: Learning How Actions Sound from Egocentric Videos, Changan Chen, Kumar Ashutosh, Rohit Girdhar, David Harwath, Kristin Grauman
299 Accept the Modality Gap: An Exploration in the Hyperbolic Space, ameera Ramasisinghe, Violetta Shevchenko, Gil Avraham, Ajanthan Thalaiyasingam
300 DiffSsal: Joint Audio and Video Learning for Diffusion Saliency Prediction, Junwen Xiong, Peng Zhang, Tao You, Chunyue Li, Wei Huang, Yuefei Zhe
301 DiPrompt: Disentangled Prompt Tuning for Multiple Latent Domain Generalization in Federated Learning, Sikai Bai, Jie Zhang, Song Guo, Shuaicheng Li, Jingcai Guo, Jun Hou, Tao Han, Xiaocheng Lu
302 Probabilistic Speech-Driven 3D Facial Motion Synthesis: New Benchmarks Methods and Applications, Karren Q. Yang, Anurag Ranjan, Jen-Hao Rick Chang, Raviteja Vemulapalli, Oncel Tuzel
303 DIEM: Decomposition-Integration Enhancing Multimodal Insights, Xinyi Jiang, Guoming Wang, Junhao Gu, Juncheng Li, Wenzhao Jiang, Rongxing Lu, Siliang Tang
304 MAFA: Managing False Negatives for Vision-Language Pre-training, Jaeseok Byun, Doohoon Kim, Taesup Moon
305 AV2AV: Direct Audio-Visual Speech to Audio-Visual Speech Translation with Unified Audio-Visual Speech Representation, Jeongsoo Choi, Se Jin Park, Minsu Kim, Yong Man Ro
306 Enhancing Multimodal Cooperation via Sample-level Modality Valuation, Yake Wei, Ruxuan Feng, Zihe Wang, Di Hu
307 Diff-BGM: A Diffusion Model for Video Background Music Generation, Szíze Li, Yiming Qin, Menghang Zheng, Xin Jin, Yang Liu
308 SaCo Loss: Sample-wise Affinity Consistency for Vision-Language Pre-training, Sitong Wu, Haoan Tu, Zhaotian Tian, Yukang Chen, Xiaojian Qi, Jiaya Jia
309 MoPE-CLIP: Structured Pruning for Efficient Vision-Language Models with Module-wise Pruning Error Metric, Haokun Lin, Haoai Bai, Zhihui Liu, Lu Hou, Muyi Sun, Linjia Song, Ying Wei, Zhenan Sun
310 Mitigating Noisy Correspondence by Geometrical Structure Consistency Learning, Zihua Zhao, Mengxi Chen, Tianjue Dai, Jiachao Yao, Bo Han, Ya Zhang, Yanfeng Wang
311 DetCLIPv3: Towards Versatile Generative Open-vocabulary Object Detection, Lewei Yao, Renjie Pi, Jianhua Han, Xiaodan Liang, Hang Xu, Wei Zhang, Zhenguo Li, Dan Xu
312 Leveraging Cross-Modal Neighbor Representation for Improved CLIP Classification, Chao Yi, Lu Ren, De-Chuan Zhan, Yan-Jia Ye
313 OmniVec2 - A Novel Transformer based Network for Large Scale Multimodal and Multitask Learning, Siddharth Srivastava, Gaurav Sharma
314 CoDi-2: In-Context Interleaved and Interactive Any-to-Any Generation, Zineng Tang, Ziyi Yang, Mahmoud Khademi, Yang Liu, Chenguang Zhu, Mohit Bansal
315 Differentiable Information Bottleneck for Deterministic Multi-view Clustering, Xiongqian Yan, Zhiqian Jin, Fengshou Han, Yangdong Ye
316 A Study of Dropout-Induced Modality Bias on Robustness

31 Multimodal Representation Learning by Alternating Unimodal Adaptation, Xiaohui Zhang, Jaehong Yoon, Mohit Bansal, Huaxiu Yao

318 View-Category Interactive Sharing Transform for Incomplete Multi-View Multi-Label Learning, hilong Ou, Zhe Xue, Yawen Li, Meiyi Liang, Yuanqiang Cai, Junjiang Wu

319 Scalable 3D Registration via Truncated Entry-wise Absolute Residuals, Tianyu Huang, Liangzu Peng, René Vidal, Yun-Hui Liu

320 Partial-to-Partial Shape Matching with Geometric Consistency, Viktoria Ehm, Maoxin Gao, Paul Roetzer, Marvin Eisenberger, Daniel Cremers, Florian Bernard

321 Towards Robust Learning to Optimize with Theoretical Guarantees, Qingshui Song, Wei Lin, Juncheng Wang, Hong Xu

322 From Varience to Variacy: Unbundling and Mitigating Gradient Variance in Differentiable Bundle Adjustment Layers, Swaminathan Gurumurthy, Karnik Ram, Bingqing Chen, Zachary Manchester, Zico Kolter

323 DIMP: Decentralized Iterative Merging-And-Training for Deep Learning Models, Nastaran Saadati, Min Pham, Nasla Saleem, Joshua R. Waite, Aditya Bala, Zhanqiang Jin, Chinnamon Hegde, Soumik Sarkar

324 Ink Dot-Oriented Differentiable Optimization for Neural Image Halftoning, Hao Jiang, Bingfeng Zhou, Yadong Mu

325 Are Conventional SNNs Really Efficient? A Perspective from Network Quantization, Guobin Shen, Dongcheng Zhao, Tenglong Li, Jindong Li, Yi Zeng

326 FedMef: Towards Memory-efficient Federated Dynamic Pruning, Hong Huang, Weiming Zhuang, Chen Chen, Linguan Lyu

327 3DMatch: Learning to Prompt Stable Diffusion Model for Semantic Matching, Xinghui Li, Jingyi Lu, Kai Han, Victor Adrian Prisacariu

328 Purified and Unified Steganographic Network, Guobiao Li, Sheng Li, Zicong Luo, Zhenxing Qian, Xinpeng Zhang

329 Learned Lossless Image Compression based on Bit Plane Slicing, Zhe Zhang, Huaerin Wang, Zhenzhen Chen, Shun Liu

330 Towards Calibrated Multi-label Deep Neural Networks, Jiacheng Cheng, Nuno Vasconcelos

331 Improving Generalization via Meta-Learning on Hard Samples, Nishtan Jain, Arun S. Suggala, Pradeep Shenoy

332 Learning with Structural Labels for Learning with Noisy Labels, Noo-ri Kim, Jin-Seop Lee, Jee-Hyong Lee

333 DiffuseMix: Label-Preserving Data Augmentation with Diffusion Models, Khawar Islam, Muhammad Sohail Danish, Muzammal Naseer, Abhilijit Das, Salman Khan, Fahad Shabbaz Khan

334 Relational Matching for Weakly Semi-Supervised Oriented Object Detection, Wenhao Wu, Hau-San Wong, Si Wu, Tianyou Zhang

335 Rethinking Transformers Pre-training for Multi-Spectral Satellite Image Detection, Mubasir Noman, Muzammal Naseer, Hisham Cholakkal, Rao Muhammad Anwer, Salman Khan, Fahad Shabbaz Khan

336 Unmixing Diffusion for Self-Supervised Hyperspectral Image Denoising, Haisun Zeng, Jiezheng Cao, Kai Zhai, Yongjun Zhang, Bingshu Wang, Yongbin Qin, Jia Wu

337 Multipolar Learning for Geospatial Vegetation Forecasting, Vitus Bonse, Claire Robin, Christain Requena-Mesa, Lazaro Alonso, Nuno Carvalhas, José Cortés, Zhihan Gao, Nara Linscheid, Mélanie Weynants, Markus Reichstein

338 Parameter Efficient Self-Supervised Geospatial Domain Adaptation, Linus Scheibenreif, Michael Mommert, Damian Borth

339 Bridging Remote Sensors with Multisensor Geospatial Foundation Models, Boran Han, Shuai Zhang, Jingxian Shi, Markus Reichstein

340 CLIP-Driven Open-Vocabulary 3D Scene Graph Generation via Cross-Modality Contrastive Learning, Lianggangxu Chen, Xuejiao Wang, Jiale Lu, Shaohui Lin, Changbo Wang, Gaoli He

341 Learnable Earth Parser: Discovering 3D Prototypes in Aerial Scans, Romain Loisaux, Elliot Vincent, Mathieu Aubry, Loïc Landrieu

342 Urban Knowledge Acquisition and Scene Reconstruction, Xiara Mei, Xingjun Wang, Daniel Seita, Simon Lacoste-Julien, Andreas Geiger

343 Geospatial Knowledge Graphs for Earth Observation, Yan Li, Danwei Deng, Zeyuan Wang, Yining Li, Zhenzhong Chen, Lingjuan Lyu

344 MiniVideo2Video: Inferring Video Content from Incomplete Video Inputs, Qingshui Song, Wei Lin, Juncheng Wang, Hong Xu

345 GeoChat: Grounded Large-Vision-Language Model for Remote Sensing, Kartik Kuckreja, Muhammad Sohail Danish, Muzammal Naseer, Abhilijit Das, Salman Khan, Fahad Shabbaz Khan

346 Parameter Efficient Self-Supervised Geospatial Domain Adaptation, Linus Scheibenreif, Michael Mommert, Damian Borth

347 Bridging Remote Sensors with Multisensor Geospatial Foundation Models, Boran Han, Shuai Zhang, Jingxian Shi, Markus Reichstein

348 CLIP-Driven Open-Vocabulary 3D Scene Graph Generation via Cross-Modality Contrastive Learning, Lianggangxu Chen, Xuejiao Wang, Jiale Lu, Shaohui Lin, Changbo Wang, Gaoli He

349 Learnable Earth Parser: Discovering 3D Prototypes in Aerial Scans, Romain Loisaux, Elliot Vincent, Mathieu Aubry, Loïc Landrieu

350 Geospatial Knowledge Graphs for Earth Observation, Yan Li, Danwei Deng, Zeyuan Wang, Yining Li, Zhenzhong Chen, Lingjuan Lyu

351 GeoChat: Grounded Large-Vision-Language Model for Remote Sensing, Kartik Kuckreja, Muhammad Sohail Danish, Muzammal Naseer, Abhilijit Das, Salman Khan, Fahad Shabbaz Khan

352 Parameter Efficient Self-Supervised Geospatial Domain Adaptation, Linus Scheibenreif, Michael Mommert, Damian Borth

353 Bridging Remote Sensors with Multisensor Geospatial Foundation Models, Boran Han, Shuai Zhang, Jingxian Shi, Markus Reichstein

354 CLIP-Driven Open-Vocabulary 3D Scene Graph Generation via Cross-Modality Contrastive Learning, Lianggangxu Chen, Xuejiao Wang, Jiale Lu, Shaohui Lin, Changbo Wang, Gaoli He

355 Learnable Earth Parser: Discovering 3D Prototypes in Aerial Scans, Romain Loisaux, Elliot Vincent, Mathieu Aubry, Loïc Landrieu

356 Geospatial Knowledge Graphs for Earth Observation, Yan Li, Danwei Deng, Zeyuan Wang, Yining Li, Zhenzhong Chen, Lingjuan Lyu

357 Parameter Efficient Self-Supervised Geospatial Domain Adaptation, Linus Scheibenreif, Michael Mommert, Damian Borth

358 Bridging Remote Sensors with Multisensor Geospatial Foundation Models, Boran Han, Shuai Zhang, Jingxian Shi, Markus Reichstein

359 CLIP-Driven Open-Vocabulary 3D Scene Graph Generation via Cross-Modality Contrastive Learning, Lianggangxu Chen, Xuejiao Wang, Jiale Lu, Shaohui Lin, Changbo Wang, Gaoli He

360 Learnable Earth Parser: Discovering 3D Prototypes in Aerial Scans, Romain Loisaux, Elliot Vincent, Mathieu Aubry, Loïc Landrieu

361 Geospatial Knowledge Graphs for Earth Observation, Yan Li, Danwei Deng, Zeyuan Wang, Yining Li, Zhenzhong Chen, Lingjuan Lyu

362 Parameter Efficient Self-Supervised Geospatial Domain Adaptation, Linus Scheibenreif, Michael Mommert, Damian Borth

363 Bridging Remote Sensors with Multisensor Geospatial Foundation Models, Boran Han, Shuai Zhang, Jingxian Shi, Markus Reichstein

364 CLIP-Driven Open-Vocabulary 3D Scene Graph Generation via Cross-Modality Contrastive Learning, Lianggangxu Chen, Xuejiao Wang, Jiale Lu, Shaohui Lin, Changbo Wang, Gaoli He

365 Learnable Earth Parser: Discovering 3D Prototypes in Aerial Scans, Romain Loisaux, Elliot Vincent, Mathieu Aubry, Loïc Landrieu

366 Geospatial Knowledge Graphs for Earth Observation, Yan Li, Danwei Deng, Zeyuan Wang, Yining Li, Zhenzhong Chen, Lingjuan Lyu

367 Parameter Efficient Self-Supervised Geospatial Domain Adaptation, Linus Scheibenreif, Michael Mommert, Damian Borth

368 Bridging Remote Sensors with Multisensor Geospatial Foundation Models, Boran Han, Shuai Zhang, Jingxian Shi, Markus Reichstein

369 CLIP-Driven Open-Vocabulary 3D Scene Graph Generation via Cross-Modality Contrastive Learning, Lianggangxu Chen, Xuejiao Wang, Jiale Lu, Shaohui Lin, Changbo Wang, Gaoli He

370 Learnable Earth Parser: Discovering 3D Prototypes in Aerial Scans, Romain Loisaux, Elliot Vincent, Mathieu Aubry, Loïc Landrieu
367 Bilateral Adaptation for Human-Object Interaction Detection with Occlusion-Robustness, Guangzhui Wang, Yangyang Guo, Zwei Xu, Mohan Kankanhalli

368 CurveCloudNet: Processing Point Clouds with 1D Structure, Colton Stearns, Alex Fu, Jiatieng Liu, Jeong Joon Park, Davis Rempe, Despoina Paschalidou, Leonidas J. Guibas

369 VCoder: Versatile Vision Encoders for Multimodal Large Language Models, Jitsh Jain, Jianwei Yang, Humphrey Shi

370 Amodal Ground Truth and Completion in the Wild, Guanqi Zhan, Chunxiao Zheng, Weidi Xie, Andrew Zisserman

371 Living Scenes: Multi-object Relocalization and Reconstruction in Changing 3D Environments, Liyuan Zhu, Shengyu Huang, Konrad Schindler, Iro Armeni

372 Single Domain Generalization for Crowd Counting, Zhuoxuan Peng, S.-H. Gary Chan

373 LTA-PCS: Learnable Task-Agnostic Point Cloud Sampling, Jiaheng Liu, Jianhao Li, Kaisiyuan Wang, Hongcheng Guo, Jiaqian Yang, Junran Peng, Ke Xu, Xianglong Liu, Jinyang Guo

374 Prompt3D: Random Prompt Assisted Weakly-Supervised 3D Object Detection, Xiaohong Zhang, Huisheng Ye, Jingwen Li, Qinyu Tang, Yuwen Guo, Jie Guo


376 Semantic Line Combination Detector, Jinwon Ko, Dongkwan Jin, Chang-Su Kim

377 From Pixels to Graphs: Open-Vocabulary Scene Graph Generation with Vision-Language Models, Rongjie Li, Songyang Zhang, Dahu Lin, Kai Chen, Xuming He

378 PanoContext-Former: Panoramic Total Scene Understanding with a Transformer, Yuan Dong, Chuan Fang, Liefeng Bo, Zilong Dong, Ping Tan

379 DiffAssemble: A Unified Graph-Diffusion Model for 2D and 3D Reassembly, Gianluca Scarpellini, Stefano Fiorini, Francesco Giuliari, Pietro Moreiro, Alessio Del Bue

380 ProMotion: Prototypes As Motion Learners, Yawen Lu, Dongfang Liu, Qifan Wang, Cheng Han, Yiming Cui, Zhwen Cao, Xueling Zhang, Yingjie Victor Chen, Heng Fan

381 HUNTER: Unsupervised Human-centric 3D Detection via Transferring Knowledge from Synthetic Instances to Real Scenes, Yichen Yao, Zimo Jiang, Yujing Sun, Zhencai Zhu, Xinge Li, Xiaohong Zhang, Huisheng Ye, Jingwen Li, Ming-Hsuan Yang

382 Rethinking the Up-Sampling Operations in CNN-based Generative Network for Generalizable Deepfake Detection, Chuangchuan Tan, Yao Zhao, Shikui Wei, Guanghua Gu, Ping Liu, Yunchao Wei

383 Shadows Don't Lie and Lines Can't Bend! Generative Models don’t know Projective Geometry. For now, Alyus Sarkar, Hanlin Mai, Amitabh Mahapatra, Svetlana Lazebnik, D.A Forsyth, Anand Bhattad

384 Text Grouping Adapter: Adapting Pre-trained Text Detector for Layout Analysis, Tiansi Bi, Xiaoyi Zhang, Zhizheng Zhang, Wenxuan Xie, Cuiling Lan, Yan Lu, Nanning Zheng

385 Groupwise Query Specialization and Quality-Aware Multi-Assignment for Transformer-based Visual Relationship Detection, Jongha Kim, IJhwon Park, Jinyoung Park, Jinyoung Kim, Sehyung Kim, Hyunwoo J. Kim

386 CoralSCOP: Segment any CORal Image on this Planet, Zhiqiang Zheng, Haixin Liang, Binh-Son Hua, Yue Him Wong, Put Ang Jr, Apple Pui Yi Chui, Sai-Kit Yeung

387 Going Beyond Multi-Task Dense Prediction with Synergy Embedding Models, Huimin Huang, Yawen Huang, Lanfen Lin, Ruofeng Tong, Yen-Wei Chen, Hao Zhang, Yuxiwei Li, Yefeng Zheng

388 Disentangled Pre-training for Human-Object Interaction Detection, Zhulong Li, Xingao Li, Changxing Ding, Xiangmin Xu

389 Osprey: Pixel Understanding with Visual Instruction Tuning, Yiqun Yuan, Wentong Li, Jian Liu, Dongqi Tang, Xinjie Luo, Chi Qin, Lei Zhang, Jianke Zhu

390 Discovering Syntactic Interaction Clues for Human-Object Interaction Detection, Jinguo Luo, Weiichong Ren, Weibo Jiang

391 Xi’ai Chen, Qiang Wang, Zhi Han, Honghai Liu

392 Flattening the Parent Bias: Hierarchical Semantic Segmentation in the Poincaré Ball, Simon Weber, Barry Zongyi, Nikola Araslanov, Daniel Cremers

393 HIKER-SGG: Hierarchical Knowledge Enhanced Robust Scene Graph Generation, Ce Zhang, Simon Stepputis, Joseph Campbell, Katia Sycara, Yaqi Xie

394 Hierarchical Intra-modal Correlation Learning for Label-free 3D Semantic Segmentation, Xin Kang, Lei Chu, Jiahao Li, Xuejin Chen, Yan Lu

395 FreePoint: Unsupervised Point Cloud Instance Segmentation, Zhikai Zhang, Jian Ding, Li Jiang, Dengxin Dai, Guisong Xia


397 MaskClustering: View Consensus based Mask Graph Clustering for Open-Vocabulary 3D Instance Segmentation, Mi Yan, Jiazhao Zhang, Yan Zhu, He Wang

398 ECoDepth: Effective Conditioning of Diffusion Models for Monocular Depth Estimation, Suraj Patni, Aradhya Agarwal, Chetan Arora

399 Physical Property Understanding from Language-Embedded Feature Fields, Albert J. Zhai, Yuan Shen, Emily Y. Chen, Gloria X. Wang, Xinfel Wang, Sheng Wang, Kaiming Yu, Shenhong Yang

400 LLM4SGG: Large Language Models for Weakly Supervised Scene Graph Generation, Kibum Kim, Kanghoon Yoon, Jaehyoe Jeon, Yeonjun In, Jinyoung Moon, Donghyun Kim, Chanyoung Park

401 DSSG: Dense Relation Transformer for an End-to-End Scene Graph Generation, Zeeshan Hayder, Xuming He

402 OTE: Exploring Accurate Scene Text Recognition Using One Token, Jianjun Xu, Yuxin Wang, Hongtao Xie, Yongdong Zhang

403 SemCity: Semantic Scene Generation with Triplane Diffusion, Jumin Lee, Sebin Lee, Changho Jo, Woobin Im, Yujiung Seon, Sung-Eui Yoon

404 Advancing Salience Ranking with Human Fixations: Dataset Models and Benchmarks, Bowen Deng, Siyang Song, Andrew P. French, Denis Schlupeck, Michael P. Pound

405 Choose What You Need: Disentangled Representation Learning for Scene Text Recognition Removal and Editing, Boqiang Zhang, Hongtao Xie, Zuan Gao, Yuxin Wang

406 Leveraging Predicate and Triplet Learning for Scene Graph Generation, Jiankai Li, Yunhong Wang, Xiefan Guo, Ruijie Yang, Weixin Li

407 Regressor-Segments Mutual Prompt Learning for Crowd Counting, Mingyue Guo, Li Yuan, Zhaoai Yan, Binhchu Chen, Yaowei Wang, Qixiang Ye

408 Learning from Observer Gaze: Zero-Shot Attention Prediction Oriented by Human-Object Interaction Recognition, Yuchen Zhou, Linkai Liu, Chao Guo

409 EGTR: Extracting Graph from Transformer for Scene Graph Generation, Jinhae Im, Jeongyeon Nam, Nokyung Park, Hyungmin Lee, Seunghyun Park

410 SG-PGM: Partial Graph Matching Network with Semantic Geometric Fusion for 3D Scene Graph Alignment and Its Downstream Tasks, Yaxu Xie, Alain Pagani, Didier Stricker

411 Open-Vocabulary Semantic Segmentation with Image Embedding Balancing, Xiangheng Shan, Dongyue Wu, Guilin Zhu, Yuanjie Shao, Nong Sang, Changxin Gao

412 Bridging the Synthetic-to-Authentic Gap: Distortion-Guided Unsupervised Domain Adaptation for Blind Image Quality Assessment, Aobo Li, Jinjian Wu, Yongxiu Liu, Leida Li


414 Robust Distillation via Untargeted and Targeted Intermediate Adversarial Samples, Junhao Dong, Piotr Koniusz, Junxi Chen, Z. Jane Wang, Yew-Soon Ong
414  Class Incremental Learning with Multi-Teacher
Distillation, Haihao Wen, Lili Pan, Yu Dai, Heqian Qiu,
Lanxiang Wang, Qinhao Wu, Hongliang Li
415  Large Language Models are Good Prompt Learners for Low-Shot
Image Classification, Zhaoheng Zheng, Jingmin Wei, Xuefeng Hu,
Haidong Zhu, Ram Nevatia
416  Consistent Prompting for Rehearsal-Free Continual
Learning, Zhanxin Gao, Jun Cen, Xiaobin Chang
417  Tuning Stable Rank Shrinkage: Aiming at the Overlooked
Structural Risk in Fine-tuning, Sicong Shen, Yang Zhou,
Bingzheng Wei, Eric I-Chao Chang, Yan Xu
418  Coherent Temporal Synthesis for Incremental Action
Segmentation, Guodong Ding, Hans Golong, Angela Yao
419  FCS: Feature Calibration and Separation for Non-Exemplar Class
Incremental Learning, Qiwei Li, Yuxin Peng, Jiahuan Zhou
420  Dell.: Direct-and-Inverse CLIP for Open-World Few-Shot
Learning, Shuai Shao, Yu Bai, Yan Wang, Baodi Liu, Yichen Gong
421  Understanding and Improving Source-free Domain Adaptation
from a Theoretical Perspective, Yu Mitsuzumi, Akisato Kimura,
Hisashi Kashima
422  Resurrecting Old Classes with New Data for Exemplar-Free
Continual Learning, Dipam Goswami, Albin Soutif-Cornerais,
Yuyang Liu, Sandesh Kamath, Bartlomiej Twardowski,
Joost van de Weijer
423  Adversarially Robust Few-shot Learning via Parameter Co-
distillation of Similarity and Class Concept Learners,
Junhao Dong, Piotr Koniusz, Junxi Chen, Xiaohua Xie, Yew-Soon Ong
424  Learning CNN on VIT: A Hybrid Model to Explicitly Class-specific
Boundaries for Domain Adaptation, Ba Hung Ngo, Nhat-Tuong
Do-Tran, Tuan-Ngoc Nguyen, Hae-Gon Jeon, Tae Jong Choi
425  Efficient Stitchable Task Adaptation, Haoyu He, Ziheng Pan, Jing
Liu, Jianfei Cai, Bohan Zhang
426  Gradient-based Parameter Selection for Efficient Fine-Tuning,
Zhi Zhang, Qizhe Zhang, Zijun Gao, Renrui Zhang, Ekaterina
Shutova, Shiji Zhou, Shanghang Zhang
427  ArGue: Attribute-Guided Prompt Tuning for Vision-Language
Models, Xinyu Tian, Shu Zou, Zhaoyuan Yang, Jing Zhang
428  Simple Semantic-Aided Few-Shot Learning, Hai Zhang, Junzhe
Xu, Shanlin Jiang, Zhenan He
429  Long-Tail Class Incremental Learning via Independent Sub-
prototype Construction, Xi Wang, Xu Yang, Jie Yin, Kun Wei,
Cheng Deng
430  Few-Shot Object Detection with Foundation Models,
Guangxing Han, Ser-Nam Lim
431  Stronger Fewer & Superior: Harnessing Vision
Foundation Models for Domain Generalized Semantic
Segmentation, Zhixiang Wei, Lin Chen, Yi Jin, Xiaoxiao Ma,
Tianle Liu, Pengyang Ling, Ben Wang, Huanan Chen, Jinjin Zheng
432  Continental Forgetting for Pre-trained Vision Models, Hongbo
Zhao, Bolin Ni, Junsong Fan, Yuxi Wang, Yuntao Chen, Gaofeng
Meng, Zhaoxiang Zheng
433  AETTA: Label-Free Accuracy Estimation for Test-Time
Adaptation, Taekyoung Lee, Sorn Chattananurak, Taesik Gong,
Sung-Ju Lee
434  Continual-MAE: Adaptive Distribution Masked Autoencoders for
Continual Test-Time Adaptation, Jiaming Liu, Ran Xu, Senqiao
Yang, Renrui Zhang, Qizhe Zhang, Zehui Chen, Yandong Guo,
Shanghang Zhang
435  LEAD: Exploring Logit Space Evolution for Model Selection,
Zixuan Hu, Xiaotong Li, Shixiang Tang, Jun Liu, Yichun Hu, Ling-Yu Duan
436  Instance-based Max-margin for Practical Few-shot
Recognition, Minghao Fu, Ke Zhu
437  Domain Gap Embeddings for Generative Dataset
Augmentation, Yining Oliver Wang, Younjoon Chung, Chen Henry Wu,
Fernando De la Torre
438  JoAPR: Cleaning the Lens of Prompt Learning for Vision-
Language Models, Yuncheng Guo, Xiaodong Gu
439  Generative Multi-modal Models are Good Class Incremental
Learners, Xusheng Cao, Haoqi Lu, Linlan Huang, Xiaolei Liu,
Ming-Ming Cheng
440  Dual Memory Networks: A Versatile Adaptation Approach for
Vision-Language Models, Yabin Zhang, Wenjie Zhu, Hui Tang,
Zhiyuan Ma, Kaiming Zhou, Lei Zhang
441  UniPT: Universal Parallel Tuning for Transfer Learning with
Efficient Parameter and Memory, Haiwen Diao, Bo Wan, Ying
Zhang, Xu Jia, Huchuan Lu, Long Chen
442  Federated Generalized Category Discovery, Nan Pu, Wenjing Li,
Xingyuan Ji, Yalan Qin, Nicu Sebe, Zhun Zhong
443  Learning from One Continuous Video Stream, João Carreira,
Michael King, Viorica Patrascu, Dilara Gokay, Catalin Ionescu,
Yi Yang, Daniel Zoran, Joseph Heyward, Carl Doersch, Yusuf Aytar,
Dima Damen, Andrew Zisserman
444  OrCo: Towards Better Generalization via Orthogonality and
Contrast for Few-Shot Class-Incremental Learning, Noor Ahmed,
Anna Kukleva, Berni Schiele
445  SDDGR: Stable Diffusion-based Deep Generative Replay for Class
Incremental Object Detection, Junsu Kim, Hoseong Cho, Jiheon
Kim, Yihaem Yimolal Tiruneh, Seungryu Baek
446  Active Domain Adaptation with False Negative Prediction for
Object Detection, Yuzuru Nakamura, Yasunori Ishii,
Takayoshi Yamashita
447  Stationary Representations: Optimally Approximating
Compatibility and Implications for Improved Model
Replacements, Niccolò Bioni, Federico Pernici, Simone Ricci,
Alberto Del Bimbo
448  Your Transferability Barrier is Fragile: Free-Lunch for Transferring
the Non-Transferable Learning, Ziming Hong, Li Shen,
Tongliang Liu
449  Transductive Zero-Shot and Few-Shot CLIP,
égléene Martin, Yushui Huang, Fereshiteh Shakeri,
Jean-Christophe Pesquet, Ismail Ben Ayed
450  Task2Box: Box Embeddings for Modeling Asymmetric Task
Relationships, Rangel Daroya, Aaron Sun, Subhransu Maji
451  Unbiased Faster R-CNN for Single-source Domain Generalized
Object Detection, Yajing Liu, Shijun Zhou, Xiyou Liu, Chunhui Hao,
Baojie Tian, Jiandong Tian
452  MetaCloak: Preventing Unauthorized Subject-driven Text-to-
image Diffusion-based Synthesis via Meta-learning,
Yixin Liu, Chenrui Fan, Yutong Dai, Xin Chen, Pan Zhou, Lichao Sun

Notes:
The Arch building is at 705 Pike Street.
The main pedestrian entrance to Arch is on the corner of 7th Avenue and Pike Street, and the Arch drop-off points are 725 Pike (private/rideshare) and 800 Convention Place (bus).

The Summit building is at 900 Pine Street, just over a block away from Arch.
The main pedestrian entrance is on the corner of 9th Avenue and Pine Street, and the drop-off location is on 9th Avenue between Pine Street and Olive Way (private/rideshare and bus).