

Weicai Ye

Curriculum Vitae

*07. March 1996, China
State Key Lab of CAD&CG
Zhejiang University (ZJU)
☎ (+86) 13032893650
✉ maikeyeweicai@gmail.com
🏠 Homepage
🐙 Github
🐦 Twitter
🔗 GoogleScholar



Education

- 2022.09– **Visiting Researcher, 3D Vision, Computer Vision and Geometry Lab, ETH Zurich, Switzerland.**
2023.03 Research Topics: Diffusion Models, SLAM and Neural Rendering. Advised by [Prof. Marc Pollefeys](#).
2018.09– **PhD, Computer Science, State Key Lab of CAD&CG, Zhejiang University (ZJU), China.**
2024.09 Research Interests: 3D Vision and Foundation Models, especially Diffusion Models, Correspondence, SLAM, (expected) Reconstruction, Rendering, and Generation. Supervised by [Prof. Hujun Bao](#) and [Prof. Guofeng Zhang](#).
2014–2018 **Bachelor of Computer Software Engineering, University of Electronic Science and Technology of China (UESTC), China.**
National Encouragement Scholarship. Ranked 3rd of 111 Students.

Publications

* denotes equal contribution, † denotes corresponding author, ‡ denotes project lead.

- 2024 Haoyi Zhu, Yating Wang, Di Huang, **Weicai Ye**, Wanli Ouyang, and Tong He. Point Cloud Matters: Rethinking the Impact of Different Observation Spaces on Robot Learning. In **NeurIPS Datasets and Benchmarks**, Under Review, 2024, **CCF-A**. **Implied that point cloud observation matters for robot learning. With point cloud as input, the agent achieved higher mean success rates and exhibited better generalization ability.**
- 2024 Xingyuan Yu*, **Weicai Ye***‡, Yuhang Ming, Jinyu Li, Hujun Bao, Zhaopeng Cui, and Guofeng Zhang†. D3FlowSLAM: Self-Supervised Dynamic SLAM with Flow Motion Decomposition and DINO Guidance. In **NeurIPS**, Under Review, 2024, **CCF-A**. **First self-supervised dynamic SLAM with Motion Decomposition and DINO Guidance, outperforming DROID-SLAM.**
- 2024 Yifan Wang, Di Huang, **Weicai Ye**†, Guofeng Zhang, Wanli Ouyang, and Tong He†. NeuRodin: A Two-stage Framework for High-Fidelity Neural Surface Reconstruction. In **NeurIPS**, Under Review, 2024, **CCF-A**. **Identified two main factors of the SDF-based approach that degrade surface quality and proposed a two-stage neural surface reconstruction framework, achieving high-fidelity surface reconstruction with fine-grained details, outperforming Neuralangelo.**
- 2024 **Weicai Ye***‡, Chenhao Ji*, Zheng Chen, Junyao Gao, Xiaoshui Huang, Song-Hai Zhang, Wanli Ouyang, Tong He†, Cairong Zhao†, and Guofeng Zhang†. DiffPano: Scalable and Consistent Text to Panorama Generation with Spherical Epipolar-Aware Diffusion. In **NeurIPS**, Under Review, 2024, **CCF-A**. **First scalable and consistent text-to-panorama generation with spherical epipolar-aware diffusion. Established large-scale panoramic video-text datasets with corresponding depth and camera poses. Achieved long-term, consistent, and diverse panoramic scene generation given unseen text and camera poses with SOTA performance.**

- 2024 **Weicai Ye**, Hao Li, Yuanyuan Gao, Yalun Dai, Junyi Chen, Nanqing Dong, Dingwen Zhang, Hujun Bao, Wanli Ouyang, Yu Qiao, Tong He†, and Guofeng Zhang†. FedSurfGS: Scalable 3D Surface Gaussian Splatting with Federated Learning. In *TPAMI*, Under Review, 2024, **CCF-A**. **First cloud-edge-device hierarchical framework with federated learning for large-scale high-fidelity surface reconstruction in a distributed manner, achieving balance between high-precision reconstruction and low-cost memory.**
- 2024 **Weicai Ye**, Xinyu Chen, Ruohao Zhan, Di Huang, Xiaoshui Huang, Haoyi Zhu, Hujun Bao, Wanli Ouyang, Tong He†, and Guofeng Zhang†. Dynamic-Aware Tracking Any Point for Structure from Motion in the Wild. In *TPAMI*, Under Review, 2024, **CCF-A**. **First propose a concise, elegant, and robust SfM pipeline with point tracking for smooth camera trajectories and dense point clouds from casual monocular videos with SOTA performance.**
- 2024 Ziyu Tang, **Weicai Ye**†, Yifan Wang, Di Huang, Hujun Bao, Tong He†, and Guofeng Zhang. ND-SDF: Learning Normal Deflection Fields for High-Fidelity Indoor Reconstruction. In *NeurIPS*, Under Review, 2024, **CCF-A**. **Proposed Normal Deflection fields to represent the angle deviation between the scene normals and the prior normals, achieving smooth surfaces with fine-grained structures, outperforming MonoSDF.**
- 2024 Yuhang Ming, Minyang Xu, Xingrui Yang, **Weicai Ye**, Weihang Wang, Yong Peng, Weichen Dai, and Wanzeng Kong. VIPeR: Visual Incremental Place Recognition with Adaptive Mining and Lifelong Learning. In *RAL*, Under Review, 2024, **Propose a novel approach for visual incremental place recognition with the ability to adapt to new environments while retaining the performance of previous environments.**
- 2024 Jiuming Liu, Ruiji Yu, Yian Wang, Yu Zheng, Tianchen Deng, **Weicai Ye**, and Hesheng Wang. Point Mamba: A Novel Point Cloud Backbone Based on State Space Model with Octree-Based Ordering Strategy. In *NeurIPS*, Under Review, 2024, **CCF-A**. **First efficient point cloud backbone with Mamba framework and achieved SOTA performance.**
- 2024 Jiuming Liu, Guangming Wang, **Weicai Ye**, Chaokang Jiang, Jinru Han, Zhe Liu, Guofeng Zhang, Dalong Du, and Hesheng Wang. DiffFlow3D: Toward Robust Uncertainty-Aware Scene Flow Estimation with Diffusion Model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, **CCF-A**. **First work of diffusion-based scene flow estimation with SOTA performance.**
- 2024 Chenxi Huang, Yuenan Hou, **Weicai Ye**, Di Huang, Xiaoshui Huang, Binbin Lin, Deng Cai, and Wanli Ouyang. NeRF-Det++: Incorporating Semantic Cues and Perspective-aware Depth Supervision for Indoor Multi-View 3D Detection. In *TCSVT*, Under Review, 2024, **Incorporating semantic cues and perspective-aware depth supervision, NeRF-Det++ outperforms NeRF-Det by +1.9% in mAP@0.25 and +3.5% in mAP@0.50 on ScanNetV2.**
- 2024 Peng Gao, Le Zhuo, Dongyang Liu, Ruoyi Du, Xu Luo, Longtian Qiu, Yuhang Zhang, Rongjie Huang, Shijie Geng, Renrui Zhang, Junlin Xie, Wenqi Shao, Zhengkai Jiang, Tianshuo Yang, **Weicai Ye**, Tong He, Jingwen He, Yu Qiao, and Hongsheng Li. Lumina-T2X: Transforming Text into Any Modality, Resolution, and Duration via Flow-based Large Diffusion Transformers. In *NeurIPS*, Under Review, 2024, **CCF-A**. **First proposed flow-based large diffusion transformers foundation model for transforming text into any modality (image, video, 3D, Audio, music, etc.), resolution, and duration.**
- 2024 Xiao Cui*, **Weicai Ye***†‡, Yifan Wang, Guofeng Zhang, Wengang Zhou, Tong He†, and Houqiang Li. StreetSurfGS: Scalable Large Scene Surface Reconstruction with Gaussian Splatting for Urban Street Scences. In *NeurIPS*, Under Review, 2024, **CCF-A**. **First photorealistic rendering and efficient high-fidelity Large Scene Surface Reconstruction for Urban Street Scenes with Free Camera Trajectories, based on PGSR, outperforming F2NeRF.**

- 2024 Yiwen Chen, Tong He, Di Huang, **Weicai Ye**, Sijin Chen, Jiaxiang Tang, Zhongang Cai, Lei Yang, Gang Yu, Guosheng Lin, and Chi Zhang. Artist-Created Mesh Generation with Autoregressive Transformers. In *NeurIPS, Under Review*, 2024, **CCF-A. First proposed a mesh generation model capable of generating artist meshes based on a reference shape, which treats mesh tokens as language tokens and employs a pre-trained shape encoder to encode the reference shape for conditional mesh generation.**
- 2024 Junyi Chen*, **Weicai Ye***^{†‡}, Yifan Wang, Danpeng Chen, Di Huang, Wanli Ouyang, Guofeng Zhang, Yu Qiao, and Tong He[†]. GigaGS: Scaling up Planar-Based 3D Gaussians for Large Scene Surface Reconstruction. In *NeurIPS, Under Review*, 2024, **CCF-A. First photorealistic rendering and efficient high-fidelity large surface reconstruction in a divide-and-conquer manner with LOD structure, based on PGSR, outperforming Neuralangelo.**
- 2024 Danpeng Chen, Hai Li, **Weicai Ye**, Yifan Wang, Weijian Xie, Shangjin Zhai, Nan Wang, Haomin Liu, Hujun Bao, and Guofeng Zhang. PGSR: Planar-based Gaussian Splatting for Efficient and High-Fidelity Surface Reconstruction. In *TVCG, Under Review*, 2024, **CCF-A. First efficient high-fidelity surface reconstruction model, outperforming Neuralangelo with faster training (ours: 1 hour vs Neuralangelo 128+ hours) and photorealistic rendering.**
- 2023 **Weicai Ye***, Xinyue Lan*, Shuo Chen, Yuhang Ming, Xingyuan Yu, Zhaopeng Cui, Hujun Bao, and Guofeng Zhang. PVO: Panoptic Visual Odometry. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, **CCF-A. First panoptic visual odometry framework to achieve comprehensive modeling of the scene motion, geometry, and panoptic segmentation information with SOTA performance.**
- 2023 **Weicai Ye***, Shuo Chen*, Chong Bao, Hujun Bao, Marc Pollefeys, Zhaopeng Cui, and Guofeng Zhang. IntrinsicNeRF: Learning Intrinsic Neural Radiance Fields for Editable Novel View Synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, **CCF-A. First introduced intrinsic decomposition into the NeRF-based rendering and performed editable novel view synthesis in room-scale scenes with SOTA performance.**
- 2022 Hailin Yu, Youji Feng, **Weicai Ye**, Mingxuan Jiang, Hujun Bao, and Guofeng Zhang. Improving feature-based visual localization by geometry-aided matching. In *Arxiv, submitted*, 2022, **As a main solution for feature matching and visual localization, integrated into OpenXRLab, cooperated with Shanghai AI Lab. ISMAR2020 extension.**
- 2022 **Weicai Ye***, Xingyuan Yu*, Xinyue Lan, Yuhang Ming, Jinyu Li, Hujun Bao, Zhaopeng Cui, and Guofeng Zhang. DeFlowSLAM: Self-Supervised Scene Motion Decomposition for Dynamic Dense SLAM. In *Arxiv, submitted*, 2022, **Proposed a novel dual-flow representation of self-supervised scene motion decomposition for dynamic dense SLAM with SOTA performance.**
- 2022 **Weicai Ye***, Xinyue Lan*, Ge Su, Hujun Bao, Zhaopeng Cui, and Guofeng Zhang. Hybrid Tracker with Pixel and Instance for Video Panoptic Segmentation. In *Arxiv, submitted*, 2022, **Achieved SOTA performance on video panoptic segmentation from two perspectives: feature space (Instance Tracker) and spatial location (Pixel Tracker).**
- 2022 Yuhang Ming, **Weicai Ye**, and Andrew Calway. iDF-SLAM: End-to-End RGB-D SLAM with Neural Implicit Mapping and Deep Feature Tracking. In *Arxiv, submitted*, 2022, **Propose a novel end-to-end RGB-D SLAM, which adopts a feature-based deep neural tracker as frontend and a NeRF-based neural implicit mapper as the backend with SOTA performance.**
- 2021 Tianxiang Zhang, Chong Bao, Hongjia Zhai, Jiazhen Xia, **Weicai Ye**, and Guofeng Zhang. ARCargo: Multi-Device Integrated Cargo Loading Management System with Augmented Reality. In *IEEE Cyber Science and Technology Congress (CyberSciTech)*, 2021, **First AR-based multi-device integrated cargo loading management system for Aerospace.**
- 2021 **Weicai Ye**, Hai Li, Tianxiang Zhang, Xiaowei Zhou, Hujun Bao, and Guofeng Zhang. SuperPlane: 3D Plane Detection and Description from a Single Image. In *IEEE Virtual Reality and 3D User Interfaces (VR)*, 2021, **CCF-A. First introduced robust plane matching in texture-less scenes and achieved SOTA performance in image-based localization.**

- 2021 Xiangyu Liu, **Weicai Ye**, Chaoran Tian, Zhaopeng Cui, Hujun Bao, and Guofeng Zhang. Coxgraph: Multi-Robot Collaborative, Globally Consistent, Online Dense Reconstruction System. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, **Best Paper Finalist Award on Safety, Security, and Rescue Robotics in memory of Motohiro Kiso**. Produced a globally consistent dense map in real-time with less transmission load.
- 2020 Hailin Yu, **Weicai Ye**, Youji Feng, Hujun Bao, and Guofeng Zhang. Learning Bipartite Graph Matching for Robust Visual Localization. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2020, **Performed robust feature matching in viewpoint and illumination changes or repetitive patterns challenges with SOTA performance**.
- 2020 Hai Li*, **Weicai Ye***, Guofeng Zhang, Sanyuan Zhang, and Hujun Bao. Saliency Guided Subdivision for Single-View Mesh Reconstruction. In *International Conference on 3D Vision (3DV)*, 2020, **First saliency guided subdivision for mesh reconstruction and achieved the trade-off between detail generation and memory consumption**.

Selected Ongoing Projects

* denotes equal contribution, † denotes corresponding author, ‡ denotes project leader.

Match Anything: A Generic Correspondence Foundation Model

- TPAMI2024 **Weicai Ye‡**, Ruohao Zhan, Hujun Bao, Wanli Ouyang, Yu Qiao, Tong He, Guofeng Zhang.
August, **First video correspondence foundation model with largest video correspondence dataset**.
Ongoing **Unified three representations of correspondence learning**: feature matching, optical flow, and point tracking with a generic video diffusion-based refinement framework: Match Anything (MAM). With a data collection loop, **built the largest internet video correspondence dataset containing over 10K hours of Citywalk videos and travel vlogs**. Incorporated 9 highly related tasks into the same framework such as stereo matching, depth, MVS, and video depth. Significantly improved **zero-shot performance** of 7 downstream applications such as homography estimation, relative pose, visual localization, object pose, SfM, SLAM, and Embodied AI (doing).

DiffusionSfM: Toward Robust and Efficient End-to-End Dense Detector-Free Structure from Motion with Cascade Diffusion-Based Refinement

- TPAMI2024 **Weicai Ye‡**, Hujun Bao, Wanli Ouyang, Yu Qiao, Tong He, Guofeng Zhang.
August, **First video diffusion-based SfM model for large scene reconstruction. Performed robust in**
Ongoing **challenging cases, such as motion blur, textureless regions, and large scenes**. Incorporating with 3D Gaussian Splatting, improved the performance of novel view synthesis (doing). Achieved **the SOTA performance** on common benchmark datasets, such as ETH3D, IMC, 1DSfM, Deblur-NeRF, and textureless SfM datasets (doing).

VideoSAM-TAP: Consistent Segmenting and Tracking Anything from Casual Videos

- TPAMI2024 **Weicai Ye‡**, Ruohao Zhan, Hujun Bao, Wanli Ouyang, Tong He, Guofeng Zhang.
August, **First proposed a unified model for consistent video segmenting anything (VideoSAM)**
Ongoing **and tracking any points (TAP)**. VideoSAM supported by TAP can effectively solve the inconsistency of SAM segmentation given consecutive frames. In reward, VideoSAM can help TAP establish accurate region matching and effectively reduce matching redundancy (doing). Achieved SOTA performance in video segmentation and point tracking (doing).

Internship Experience

General 3D Intern, Shanghai AI Lab. Shanghai, China

Oct2023-Present ***Match Anything (Correspondence Foundation Model) & DiffusionSfM & InternVerse (Reconstruction Foundation Models include SurfelGS & FedSurfGS & GigaGS & StreetSurfGS & InvrenderGS & NeuRodin) & DiffPano (Text to Multi-view Panorama Generation) & MAIL (Embodied Foundation Model for Imitation Learning), etc. Working with Dr. Tong He, Prof. Wanli Ouyang and Prof. Yu Qiao.***

As the first author/corresponding author/project lead, established the largest internet video correspondence, reconstruction, video depth, and panoramic datasets. Proposed correspondence, reconstruction, video depth, panorama generation, and embodied foundation models (doing). Achieved SOTA zero-shot performance in several tasks and benchmarks. Mentoring 10+ junior researchers at Shanghai AI Lab.

Awards & Honors

- 2022 ***State Scholarship Fund from China Scholarship Council.***
- 2022 ***Academic Rising Star at Zhejiang University.***
- 2021 ***Best Paper Finalist Award on Safety, Security, and Rescue Robotics in memory of Motohiro Kiso.*** in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).
- 2019 ***Zhijun He Outstanding Scholarship.***
- 2019 ***Chiang Chen Industrial Charity Foundation Grant.***
- 2018 ***Champion*** of 2018 Cloudwalk Headcount Challenge with 31,500¥ Bonus.
- 2017 ***National Encouragement Scholarship.***
- 2017 ***Meritorious Winner*** in Mathematical Contest Modeling.
- 2016 ***First Prize*** in Sichuan Province Contest District in China Undergraduate Mathematical Contest in Modeling.

Skills

Programming Python, Pytorch, C++, CUDA, LATEX
Management Mentored 20+ Junior Researchers at Shanghai AI Lab, ETH, TUM, CUHK, NTU, THU, ZJU, SJTU, FDU, USTC, NWPU, and TJU.
Language Chinese (Native), English (Fluent)

Services

Conference reviewer: ICML, NeurIPS, ICLR, CVPR, ECCV, ICRA, IROS, VR, ISMAR, AAAI
Journal reviewer: TIP, RAL, CVM

Teaching Assistant

Spring, 2021: Computational Photograph

References

Prof. Guofeng Zhang

Full Professor, College of Computer Science
State Key Lab of CAD&CG, Zhejiang University
☎ + (86) 571-88208701
✉ zhangguofeng@zju.edu.cn

Prof. Hujun Bao

Cheung Kong Professor, College of Computer Science
State Key Lab of CAD&CG, Zhejiang University
☎ + (86) 13505813153
✉ baohujun@zju.edu.cn

Dr. Tong He

Research Scientist
Shanghai AI Laboratory
✉ tonghe90@gmail.com

Prof. Wanli Ouyang

Professor, Chinese University of Hong Kong
Shanghai AI Laboratory
✉ ouyangwanli@pjlab.org.cn