Yiyao Zhu

yzhucp@connect.ust.hk Google Scholar zyy721.github.io

Education

Hong Kong University of Science and Technology (HKUST) **Hong Kong**
Ph.D. in Electronic and Computer Engineering *09.2020 - 11.2026*
Supervisor: Prof. Shaojie SHEN, Co-Supervisor: Prof. Michael Yu WANG

Zhejiang University (ZJU) **Hangzhou**
B.Eng. in Electrical Engineering & Automation *09.2014 - 06.2018*
GPA: 3.86/4

Research Profile

My research focuses on learning scalable **3D/4D representations and predictive world models** from large-scale visual data with limited manual annotations. I combine geometric consistency, foundation-model supervision, and latent generative modeling to learn spatiotemporal representations for perception, prediction and decision-making in autonomous systems. I am particularly interested in extending these methods toward 3D world models for embodied perception, physical interaction, and robotic manipulation.
Research Areas: 3D/4D Vision, World Models, Self-Supervised Representation Learning, Embodied AI

Experience

Robotics Institute, HKUST **Hong Kong**
Ph.D. Candidate (3D World Models), Supervisor: Prof. Shaojie SHEN and Prof. Michael Yu WANG. *09.2020 - Present*
- Developed Gaussian-centric latent world models for predictive pre-training. [DLWM, CVPR'26]
- Developed query-based splatting pre-training for sparse perception models. [SQS, NeurIPS'25 Spotlight]
- Developed vision-centric pre-training for dense voxel representations. [VisionPAD, CVPR'25]
- Proposed bi-level future fusion for interactive trajectory prediction. [BiFF, ICCV'23]
- Leveraged pseudo-labels from 2D foundation models for annotation-efficient 3D representation learning.

The Chinese University of Hong Kong, Shenzhen **Shenzhen**
Research Engineer (Robot Manipulation), Supervisor: Prof. Yuquan WANG and Dr. Yongquan CHEN. *09.2018 - 06.2020*
- Contributed to a warehouse mobile manipulation system with perception and suction-based picking. [Project]
- Developed real-time QP-based task-priority control for a UR5 operating in dynamic environments. [Project]
- Developed constrained Bayesian optimization for sample-efficient, safe task-priority control. [Paper]

Publications & Manuscripts

[1] VisionPAD++: Vision-Centric Pre-training with Foundation Models for Autonomous Driving
Yiyao Zhu (first author)
Improves 3D/4D representations with 2D foundation models and world models.
Manuscript in preparation.

[2] DLWM: Dual Latent World Models enable Holistic Gaussian-centric Pre-training in Autonomous Driving
Yiyao Zhu, Ying Xue, Haiming Zhang, Guangfeng Jiang, Wending Zhou, Xu Yan, Jiantao Gao, Yingjie Cai, Bingbing Liu, Zhen Li, Shaojie Shen
Learns predictive Gaussian-centric representations by modeling future scene dynamics in latent space.
Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR 2026**), [Paper]

[3] WPT: World-to-Policy Transfer via Online World Model Distillation
Guangfeng Jiang, Yueru Luo, Jun Liu, Yi Huang, **Yiyao Zhu**, Zhan Qu, Dave Zhenyu Chen, Bingbing Liu, Xu Yan
Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR 2026**), [Paper]

[4] Vision-Language-Action Models for Autonomous Driving: Past, Present, and Future
Tianshuai Hu*, Xiaolu Liu*, Song Wang*, **Yiyao Zhu***, Ao Liang*, Lingdong Kong*, Guoyang Zhao, Zeying Gong, Jun Cen, Zhiyu Huang, Xiaoshuai Hao, Linfeng Li, Hang Song, Xiangtai Li, Jun Ma, Shaojie Shen, Jianke Zhu, Dacheng Tao, Ziwei Liu, Junwei Liang
IEEE Transactions on Pattern Analysis and Machine Intelligence (**TPAMI 2026**), Major Revision, [Paper]

[5] LLM-World: Enhancing LLM-based 4D Occupancy Forecasting with Video Pre-training and One-Step Generation
Ying Xue*, **Yiyao Zhu***, Pengyu Yin, Wending Zhou, Xu Yan, Jinliang Ding, Shaojie Shen, Shuguang Cui, Zhen Li
IEEE Transactions on Neural Networks and Learning Systems (**TNNLS 2026**), Under Review, [Paper]

[6] SQS: Enhancing Sparse Perception Models via Query-based Splatting in Autonomous Driving
Haiming Zhang*, **Yiyao Zhu***, Wending Zhou, Xu Yan, Yingjie Cai, Bingbing Liu, Shuguang Cui, Zhen Li
Introduces query-based splatting for pre-training sparse perception models.
Advances in Neural Information Processing Systems (**NeurIPS 2025 Spotlight, top 3.1% of 21,575**), [Paper]

[7] VisionPAD: A Vision-Centric Pre-training Paradigm for Autonomous Driving
Haiming Zhang*, Wending Zhou*, **Yiyao Zhu***, Xu Yan, Jiantao Gao, Dongfeng Bai, Yingjie Cai, Bingbing Liu, Shuguang Cui, Zhen Li
Develops a vision-centric pre-training paradigm for autonomous driving.
Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR 2025**), [Paper]

[8] Two-Stage LVLM System: 1st Place Solution for ECCV 2024 Corner Case Scene Understanding Challenge
Ying Xue, Haiming Zhang, **Yiyao Zhu**, Wending Zhou, Shuguang Cui, Zhen Li
ECCV 2024 Workshop on Multimodal Perception and Comprehension of Corner Cases in Autonomous Driving (**ECCVW 2024, Ranked 1st among 43 teams worldwide**), [Paper]

[9] BiFF: Bi-level Future Fusion with Polyline-based Coordinate for Interactive Trajectory Prediction
Yiyao Zhu, Di Luan, Shaojie Shen
Explicitly models future agent interaction for trajectory prediction.
Proceedings of the IEEE/CVF International Conference on Computer Vision (**ICCV 2023**), [Paper]

[10] Knowledge-guided Two-stage Memetic Search for the Pickup and Delivery Traveling Salesman Problem with FIFO Loading
Yiyao Zhu, Yongquan Chen, Zhang-Hua Fu
Improves upper bounds for the pickup-and-delivery traveling salesman problem (TSP) with FIFO loading.
Knowledge-Based Systems, 2022, [Paper]

[11] Sample-efficient learning of soft priorities for safe control with constrained Bayesian optimization
Jian Li, **Yiyao Zhu**, Langcheng Huo, Yongquan Chen
Developed Bayesian optimization for sample-efficient learning of soft task priorities in robot control.
IEEE International Conference on Robotic Computing (**IRC 2020**), [Paper]

* indicates equal contribution.

Awards & Honours

- Research Travel Grant, HKUST *2023, 2026*
- Best Teaching Assistant Award 2020-21, ECE, HKUST *06.2022*
- Postgraduate Scholarship, HKUST *09.2020*
- Finalist, 2018 JD.com Robotics Challenge (Team Prize: CNY 100,000) [Demo] *12.2018*
- The First-Class Scholarship for Excellence in Research and Innovation, ZJU *10.2017*
- Meritorious Winner of American Mathematical Contest In Modeling (MCM top 8%) *04.2017*

Invited Talks

- 3dcver Online [record]: Self-supervised Pre-training for 3D Representations in AD
- AI TIME Online [record]: DLWM: Dual Latent World Models enable Holistic Gaussian-centric Pre-training in AD

Teaching

- Teaching Assistant: ELEC1030 - The Rise of Autonomous Robots, HKUST, Spring 2021
  Instructor: Prof. Michael Yu WANG
- Teaching Assistant: ELEC4240 - Deep Learning in Computer Vision, HKUST, Fall 2021
  Instructor: Prof. Qifeng CHEN

Academic Service

- Reviewer: CVPR, ECCV, NeurIPS, ICLR, and CoRL.