Yuzhi Huang

Ph.D. Student @ mmlab@SIGS, Tsinghua SIGS · Embodied AI / 4D World Modeling / Multimodal Reasoning
yz_huang13@163.com | yu2hi13.github.io | Google Scholar | github.com/yu2hi13

Research Summary

First-year Ph.D. student at Tsinghua Shenzhen International Graduate School (mmlab@SIGS), co-advised by Prof. Zhi Wang and Dr. Jingyan Jiang. My research lies at the intersection of **Embodied AI**, **4D World Modeling**, **Multimodal Dynamic Reasoning**, and **Vision Foundation Models**. I build embodied agents that perceive, remember, reason, and act in dynamic physical environments, toward open-world embodied intelligence.

- **Embodied AI & Robotic Agents:** Lead **Mind2Act World [Tech Report 2026]**, a diagnostic benchmark that separates embodied reasoning, physical execution, and their closed-loop composition across 30 manipulation tasks, and develop **RoboStream [ECCV 2026]** and **ChainVLA** for long-horizon manipulation with persistent spatio-temporal memory and a unified execution state.
- **4D World Modeling:** Build physical-scale multimodal 4D modeling from real-world video in **DynamicVerse [NeurIPS 2025]**, and model evolving 4D worlds in latent space in **Pros4D**.
- **Multimodal Dynamic Reasoning:** Study how MLLMs perceive, track, and reason about dynamics in the physical 4D world with **Thinking in Dynamics [CVPR 2026]** *(oral at the CVPR 2026 DataMFM Workshop)* and **DynTrace [ACM MM 2026]**.
- **Robust Perception with Foundation Models:** Track-centric video anomaly understanding in **TAO [CVPR 2025]** and **TAU-Bench**, and ambiguity-aware segmentation with SAM in **ASAM [NeurIPS 2024]** and **P^{2}SAM [ACM MM 2024]**.

Education

Tsinghua Shenzhen International Graduate School, Tsinghua University — Shenzhen, China
Ph.D. Student, mmlab@SIGS — *2026 – Present*

- Advisors: Prof. Zhi Wang; Dr. Jingyan Jiang.
- Member of the Self-Evolving Embodied Agents Group.
- Research: Embodied AI, 4D World Modeling, Multimodal Dynamic Reasoning, Vision Foundation Models.

Xiamen University — Xiamen, China
Master's Degree, School of Informatics — *2023 – 2026*

- Advisors: Prof. Xinghao Ding; Prof. Yue Huang. Member of the SmartDSP Lab.

Fujian Normal University — Fuzhou, China
Bachelor's Degree — *2019 – 2023*

Research Experience

Velabot AI — Shenzhen, China
Core Algorithm Member (Intern) — *Oct. 2025 – Present*

- Embodied AI startup incubated by our research group at mmlab@SIGS.

Selected Publications

Summary: 14 papers and preprints, including 7 published at CVPR, ECCV, NeurIPS, and ACM MM. First or co-first author on **9** and corresponding author or project lead on **6**.

** Equal Contribution. † Corresponding Author or Project Lead.*

First-Author & Co-First-Author Publications

[Tech Report 2026] Mind2Act World: A Diagnostic Benchmark for Embodied Mind, Physical Action, and Mind-Action Composition in Robotic Manipulation
Yuzhi Huang*†, Jie Wu*†, Fanding Huang*, Junchao Lin*, Weichen Zhang*, Ziyi Xiong*, Yang Li*, Zhengwei Xie*, Zhimin Wang*, Zhen Liu*, Jiazhen Huang*, *et al.*, Cheng Chi†, Jingyan Jiang†, Zhi Wang†.

[ECCV 2026] RoboStream: Weaving Spatio-Temporal Reasoning with Memory in Vision-Language Models for Robotics
Yuzhi Huang*†, Jie Wu*, Weijue Bu, Ziyi Xiong, Gaoyang Jiang, Ye Li, Kangye Ji, Shuzhao Xie, Yue Huang, Chenglei Wu, Jingyan Jiang, Zhi Wang.

[ACM MM 2026] DynTrace: Tracking Dynamic Object Evidence for 4D Spatio-Temporal Reasoning in MLLMs
Rongxin Gao*, Yuzhi Huang*†, Dongxuan Liu*, Chu Li, Zhenye Wang, Jie Wu, Shuzhao Xie, Jingyan Jiang, Xinghao Ding, Xiaotong Tu, Yue Huang.

[CVPR 2026] Thinking in Dynamics: How Multimodal Large Language Models Perceive, Track, and Reason Dynamics in Physical 4D World
Yuzhi Huang*†, Kairun Wen*, Rongxin Gao*, Dongxuan Liu, Yibin Lou, Jie Wu, Jing Xu, Jian Zhang, Zheng Yang, Yunlong Lin, Chenxin Li, Panwang Pan, Junbin Lu, Jingyan Jiang, Xinghao Ding, Yue Huang, Zhi Wang.
Also an oral presentation at the CVPR 2026 DataMFM Workshop and accepted to the CVPR 2026 4D Vision Workshop.

[Preprint 2026] ChainVLA: Chaining Vision-Language-Action Queries through a Unified Execution State for Long-Horizon Manipulation
Yuzhi Huang*, Weijue Bu*, Ziyi Xiong, Jie Wu, Fanding Huang, Jingyan Jiang, Zhi Wang.

[NeurIPS 2025] DynamicVerse: Physically-Aware Multimodal Modeling for Dynamic 4D Worlds
Kairun Wen*, Yuzhi Huang*, Runyu Chen, Hui Zheng, Yunlong Lin, Panwang Pan, Chenxin Li, Wenyan Cong, Jian Zhang, Junbin Lu, Chenguo Lin, Dilin Wang, Zhicheng Yan, Hongyu Xu, Justin Theiss, Yue Huang, Xinghao Ding, Rakesh Ranjan, Zhiwen Fan.

[CVPR 2025] Track Any Anomalous Object: A Granular Video Anomaly Detection Pipeline
Yuzhi Huang*, Chenxin Li*, Haitao Zhang, Zixu Lin, Yunlong Lin, Hengyu Liu, Wuyang Li, Xinyu Liu, Jiechao Gao, Yue Huang, Xinghao Ding, Yixuan Yuan.

[NeurIPS 2024] Flaws can be Applause: Unleashing Potential of Segmenting Ambiguous Objects in SAM
Chenxin Li*, Yuzhi Huang*, Wuyang Li, Hengyu Liu, Xinyu Liu, Qing Xu, Zhen Chen, Yue Huang, Yixuan Yuan.

[ACM MM 2024] P^{2}SAM: Probabilistically Prompted SAMs Are Efficient Segmentators for Ambiguous Medical Images
Yuzhi Huang*, Chenxin Li*, Zixu Lin, Hengyu Liu, Haote Xu, Yifan Liu, Yue Huang, Xinghao Ding, Yixuan Yuan.

Corresponding-Author Publications

[Preprint 2026] Pros4D: Modeling Evolving 4D Worlds in Latent Space for Dynamic Spatial Reasoning
Rongxin Gao, Dongxuan Liu, Chu Li, Chengyang Ji, Jialing Huang, Runyu Chen, Kairun Wen, Chenxin Li, Jingrui Ye, Haosong Zhen, Xinghao Ding†, Yue Huang, Yuzhi Huang†, Xiaotong Tu, Jingyan Jiang.

[Preprint 2026] TAU-Bench: From Anomaly Instance Tracking to Fine-Grained Video Anomaly Understanding
Kepeng Yang*, Dongxuan Liu*, Rongxin Gao*, Zixin Su, Rui Wu, Shuzhao Xie, Chenxin Li, Panwang Pan, Yuzhi Huang†, Yue Huang†, Jingyan Jiang.

Co-Authored Publications

[Preprint 2026] RobotWorld: Benchmarking Multimodal Agents for Robot Use Across Diverse Tasks and Embodiments
Zhiqin Yang*, Chenxin Li*†, Xiaomeng Hu*†, Yibin Liu, Weidong Huang, Jiankai Sun, Haitao Li, Zijian Wu, Yuzhi Huang, Fanding Huang, *et al.*, Tsung-Yi Ho, Yike Guo.

[Preprint 2026] ChunkTrust: Adapting Execution Horizons for Robot Policies with Action-Expert Evidence
Fanding Huang*, Jingyan Jiang*, Shifeng Bao*, Mingkang Pu, Shiwei Li, Jing Xu, Shijia Xu, Guanbo Huang, Chenghao Gu, Yuzhi Huang, Chenxin Li, Faisal Nadeem Khan, Huan Yang, Yan Wang, Cheng Chi†, Zhi Wang†.

[Preprint] Zip-VGGT: Object-Centric Spatiotemporal KV Compression for Streaming Vision Transformers
Kairun Wen†, Runyu Chen, Wenyan Cong, Peiwei Lin, Tao Lu, Lihan Jiang, Weiguang Zhao, Junting Dong, Yunlong Lin, Yuzhi Huang, Xinghao Ding, Hongsheng Li, Linning Xu, Mulin Yu.

Honors & Academic Service

Honors & Awards

- **CAST Young Science and Technology Talent Cultivation Program (Doctoral Track)**, 2026
- **Tsinghua Future Scholar Scholarship**, four recipients from Tsinghua SIGS, 2026
- National Scholarship, 2026
- Outstanding Graduate, 2026
- Huang Xilie University-Level Scholarship, 2025
- University Level Outstanding Student Award, 2024
- Meritorious Winner, Mathematical Contest in Modeling (MCM), 2024

Academic Service

- **Program Committee Member:** AAAI (2026).
- **Conference Reviewer:** NeurIPS (2024, 2025), ICLR (2025), ICML (2025).