Yuanqi Yao

Website | Google Scholar | Email | Tel

RESEARCH INTERESTS

My research interests focus on Embodied AI and 3D Computer Vision, particularly vision-language-action models.

EDUCATION

Harbin Institute of Technology   **Harbin, China**
Master of Science in Computer Science and Technology, Assessment: 90.0/100.0   *Sep 2022 - Feb 2025*
Harbin Institute of Technology   **Harbin, China**
Bachelor of Science in Computer Science and Technology, Assessment: 87.4/100.0   *Sep 2018 - Jun 2022*

EXPERIENCE

INSAIT | Researcher, Advisors: **Prof. Luc Van Gool**   **Sofia, Bulgaria**
Leading INSAIT-Toyota joint project on multi-robot collaboration. (**2nd** place at **NeurIPS 2025 Workshop**)   *Sep 2025 - Present*

Shanghai AI Laboratory | Embodied AI Intern, Advisor: **Xuelong Li**   **Shanghai, China**
Leading research on VLA models and lifelong robot learning. (**CVPR 2025**, **RSS 2025 Spotlight**)   *Nov 2023 - Jul 2025*

Baidu | Computer Vision Intern, Advisor: **Jingdong Wang**   **Beijing, China**
Developed BEV hierarchical decoder for Jiyue Robocar. (**10%** complexity reduction, **5.92%** mAP improvement)   *Aug 2023 - Nov 2023*

Lenovo Research | Computer Vision Research Intern   **Beijing, China**
Leading research on event detection for autonomous driving. (**1st** and **2nd** place at **ICCV 2023 Workshop**)   *Jun 2023 - Aug 2023*

NIO | Autonomous Driving Perception Intern   **Beijing, China**
Optimized NIO's TorchPilot framework with parallel processing. (**10x** deployment speed improvement)   *Jun 2022 - Oct 2022*

SELECTED PUBLICATION

1. **Yuanqi Yao**, Siao Liu, Zhigang Wang, Bin Zhao, Yan Ding, Dong Wang, Xuelong Li. "Think Small, Act Big: Primitive Prompt Learning for Lifelong Robot Manipulation". **(CVPR 2025 Accepted)**.
   - We propose PPL (Primitive Prompt Learning), a two-stage framework for lifelong robot learning that captures shared primitives via motion-aware prompt learning and enables knowledge transfer between old and new skills.

2. **Yuanqi Yao**, Gang Wu, Kui Jiang, Siao Liu, Jian Kuai, Xianming Liu, Junjun Jiang."Improving Domain Generalization in Self-Supervised Monocular Depth Estimation via Stabilized Adversarial Training". (**ECCV 2024 Accepted**).
   - We propose SCAT, a novel framework for self-supervised Depth Estimation that introduces a stabilized adversarial training approach, significantly enhancing model stability and generalization, and achieves SotA performance across multiple benchmarks.

3. Delin Qu*,Haoming Song*,Qizhi Chen*,**Yuanqi Yao**, Xinyi Ye, Jiayuan Gu, Bin Zhao, Dong Wang."SpatialVLA: Exploring Spatial Representations for Visual-Language-Action Models". **(RSS 2025 Spotlight Accepted)**.
   - We present SpatialVLA, a spatial-enhanced vision-language-action model trained on 1.1M real robot episodes with 3D position encoding and adaptive spatial grids. It achieves SOTA performance with better generalization and faster inference speed.

4. Pihai Sun, Junjun Jiang†, **Yuanqi Yao**, Youyu Chen, Wenbo Zhao, Kui Jiang, Xianming Liu."FUSE: Label-Free Image-Event Joint Monocular Depth Estimation via Frequency-Decoupled Alignment and Degradation-Robust Fusion". **(IROS 2025 Accepted)**.
   - We propose FUSE, a self-supervised image-event depth estimation framework that achieves SotA performance with 14% and 24.9% improvements on MVSEC and DENSE datasets, demonstrating zero-shot adaptability to lighting and motion blur.

5. Haoming Song*,Delin Qu*, **Yuanqi Yao**, Qizhi Chen, Qi Lv, Yiwen Tang, Modi Shi, Guanghui Ren, Maoqing Yao, Bin Zhao, Dong Wang, Xuelong Li."Hume: Introducing System-2 Thinking in Visual-Language-Action Model". **(Under Review)**.
   - We present Hume, a dual-system Vision-Language-Action (VLA) model that integrates System-2 thinking with value-guided action selection and cascaded action denoising, enabling deep reasoning and precise control for humanoid robots.

AWARDS AND HONOURS

1. **2nd** place at Multi-Agent Embodied Intelligence Challenge, Control Track (**NeurIPS 2025 Workshop**).
2. **2nd** place at AIM 2024 ECCV Depth Upsampling Challenge (**ECCV 2024 Workshop**).
3. **1st** place at ICCV 2023 The ROAD++ Challenge, Spatiotemporal Agent Detection Track (**ICCV 2023 Workshop**).
4. **2nd** place at ICCV 2023 The ROAD++ Challenge, Spatiotemporal Road Event Detection Track (**ICCV 2023 Workshop**).
5. **3rd** place at ICRA 2023 The RoboDepth Challenge, Self-supervised Depth Estimation Track (**ICRA 2023 Workshop**).
6. **1st** Scholarship for Postgraduate Students, 2023-2024.
7. The People's Scholarship in China, 2018-2020.
8. **National Second Prize (Top 1%)**, National University IoT Design Competition, **Huawei Cup**, 2019.