ZIHAO WANG

Email: zhwang@stu.pku.edu.cn · Homepage: https://zhwang4ai.github.io

INTERESTS

I work on building open-ended embodied agents with multi-task skills, including task planning, decision-making, and visual localization. In particular, I am interested in building and leveraging large pre-trained Foundation Models to improve the generalization of agent capabilities. My long-horizon goal is to build interactive open-world agents capable of understanding human instructions and executing tasks with human-like planning and reasoning.

Recently, we have developed a series of open-world multi-task agents, including **OmniJARVIS** (pretrained end-to-end Vision-Language-Action models with self-supervised quantified behavior tokenizer), **JARVIS-1** (self-improving with multimodal memory), **DEPS** (interactive long-horizon planning agent), **RAT** (tool-use agent with retrieval-augmented thought), **GROOT** (self-supervised vision-based multitask policy), and **ProAgent** (collaborating agents).

EDUCATION

Peking University **Beijing, China**

Ph.D. Student in Artificial Intelligence (AI) *2022.09 - Present*

- *Advisor*: Professor Yitao Liang

Beijing Institute of Technology **Beijing, China**

M.S. in Control Science and Technology *2019 - 2022*

- *Advisor*: Professor Zhen Li

Beijing Institute of Technology **Beijing, China**

B.S. in Automation *2015 - 2019*

PROFESSIONAL EXPERIENCE

Research Intern, AI Lab in Ailibaba Inc, Beijing, CHINA *2021.06 - 2021.08*

PUBLICATIONS

- **Reasoning and Planning in LLM Agents**

  [1] Zihao Wang, Shaofei Cai, Anji Liu, Yonggang Jin, Jinbing Hou, Bowei Zhang, Haowei Lin, Zhaofeng He, Zilong Zheng, Yaodong Yang, Xiaojian Ma, Yitao Liang. JARVIS-1: Open-world multi-task agents with memory-augmented multimodal language models.
  T-PAMI, arXiv:2311.05997, 2023. [Page]

  [2] Zihao Wang, Shaofei Cai, Guanzhou Chen, Anji Liu, Xiaojian Ma, Yitao Liang. Describe, Explain, Plan and Select: Interactive Planning with Large Language Models Enables Open-World Multi-Task Agents.
  NeurIPS, 2023; also appeared on ICML 2023 TEACH Workshop Best Paper. [Code]

  [3] Zihao Wang, Anji Liu, Haowei Lin, Jiaqi Li, Xiaojian Ma, Yitao Liang. Retrieval Augmented Thoughts Elicit Context-Aware Reasoning in Long-Horizon Generation.
  arXiv preprint arXiv:2403.05313, 2024. [Page]

[4] Ceyao Zhang, Kaijie Yang, Siyi Hu, Zihao Wang, Guanghe Li, Yihang Sun, Cheng Zhang, Zhaowei Zhang, Anji Liu, Song-Chun Zhu, Xiaojun Chang, Junge Zhang, Feng Yin, Yitao Liang, Yaodong Yang. ProAgent: Building Proactive Cooperative Agents with Large Language Models. AAAI 2024 (**Oral**). [Page]

- **Foundation Model for Decision-making**

[5] Zihao Wang, Shaofei Cai, Zhancun Mu, Haowei Lin, Ceyao Zhang, Xuejie Liu, Qing Li, Anji Liu, Xiaojian Ma, Yitao Liang. OmniJARVIS: Unified Vision-Language-Action Tokenization Enables Open-World Instruction Following Agents. NeurIPS 2024. [Page]

[6] Shaofei Cai, Zihao Wang, Xiaojian Ma, Anji Liu, Yitao Liang. Open-world multi-task control through goal-aware representation learning and adaptive horizon prediction. CVPR, 2023. [Code]

[7] Shaofei Cai, Bowei Zhang, Zihao Wang, Xiaojian Ma, Anji Liu, Yitao Liang. GROOT: Learning to Follow Instructions by Watching Gameplay Videos. ICLR 2024 (**Spotlight**). [Page]

[8] Shaofei Cai, Zihao Wang, Kewei Lian, Zhancun Mu, Xiaojian Ma, Anji Liu, Yitao Liang. ROCKET-1: Master Open-World Interaction with Visual-Temporal Context Prompting. CVPR 2025.

- **Visual Geometry and SLAM**

[9] Zihao Wang, Chunxu Wu, Yifei Yang, Zhen Li. Learning Transformation-Predictive Representations for Detection and Description of Local Features. CVPR, 2023.

[10] Zihao Wang, Zhen Li, Xueyi Li, Wenjie Chen, Xiangdong Liu. Graph-Based Contrastive Learning for Description and Detection of Local Features. IEEE Trans. Neural Netw. Learn. Syst. (TNNLS) 2022.

[11] Zihao Wang, Xueyi Li, Zhen Li Local Representation is NOT Enough: Soft Point-wise Transformer for descriptor and Detector of Local Features. IJCAI 2021.

- **Benchmarks and Others**

[12] Haowei Lin, Baizhou Huang, Haotian Ye, Qinyu Chen, Zihao Wang, Sujian Li, Jianzhu Ma, Xiaojun Wan, James Zou, Yitao Liang. Selecting Large Language Model to Fine-tune via Rectified Scaling Law. ICML 2024.

[13] Haowei Lin, Zihao Wang, Jianzhu Ma, Yitao Liang. MCU: A Task-centric Framework for Open-ended Agent Evaluation in Minecraft. ALOE Workshop at NeurIPS 2023.

[14] Yuheng Cheng, Ceyao Zhang, Zhengwen Zhang, Xiangrui Meng, Sirui Hong, Wenhao Li, Zihao Wang, Zekai Wang, Feng Yin, Junhua Zhao, Xiuqiang He. Exploring large language model based intelligent agents: Definitions, methods, and prospects. *arXiv:2401.03428, 2024.*

[15] Haotian Zhang, Junting Zhou, Haowei Lin, Hang Ye, Jianhua Zhu, Zihao Wang, Liangcai Gao, Yizhou Wang, Yitao Liang. CLoG: Benchmarking Continual Learning of Image Generation Models. *arXiv preprint arXiv:2406.04584, 2024.* [Code]

AWARDS

- **National Scholarship**, CHINA, *2021*
- **Outstanding Graduate** (top 1%), Beijing, CHINA, *2019*

- **Ranking 2nd** on ECCV 2024 Multimodal Perception and Comprehension of Corner Cases in Autonomous Driving Challenges.
- **Special Prize** on Autonomy in International Micro-unmanned Aerial Vehicle Competition (IMAV), Melbourne, Australia, *2018*