Shuanghao Bai

Tel: +86-199-4940-8829 | Email: baishuanghao@stu.xjtu.edu.cn

Homepage | Google Scholar | Github

Xi'an, Shannxi - China

EDUCATION

- **Xi'an Jiaotong University** *Sept. 2022 - Jun. 2027*
  Ph.D. candidate. of Control Science and Technology Xi'an, China
  - Advisor: Badong Chen
  - Research Interests: Generalization of Machine Learning, Robotics.

- **Westlake University** *Sept. 2024 - Jun. 2025*
  Visiting Student in MiLAB Hangzhou, China
  - Advisor: Donglin Wang
  - Research Interests: Robotics.

- **Chongqing University** *Sept. 2018 - Jun. 2022*
  Bachelor of Automation Chongqing, China
  - Advisor: Min Zhao
  - GPA: 3.68/4.00 (Top 5%)

PROJECTS

- **Cloud-Edge-Device Robot Platform** *Sept. 2022 - Dec. 2025*
  Basic theories and key technologies of cloud-edge-device integrated service robot cloud-brain platform
  - Mainly focuses on generalization tasks in computer vision, addressing challenges posed by limited data and significant distribution shifts between training and test domains. (1) Proposed a simple yet effective method that significantly improves generalization in few-shot settings by incorporating multilayer perceptrons during pretraining [C.3]; (2) Unified and evaluated prompt tuning techniques on the vision-language model CLIP for unsupervised domain adaptation, and proposed a prompt-based method to mitigate domain discrepancies [C.2]; (3) Developed a generative prompt learning method for domain generalization that leverages CLIP and CGANs to learn and transfer domain-specific prompts to unseen domains [C.1].
  - Also works on generalization in embodied intelligence, with a focus on robot manipulation. (1) Analyzed representation redundancy in embodied datasets using mutual information, and introduced an information bottleneck approach to enhance generalization performance [C.5].

- **Robotic Arm Platform (Project Applicant and Leader: Shuanghao Bai)** *Jan. 2024 - Dec. 2025*
  Robotic arm platform technology and application based on visual language action model
  - Developed vision-language-action (VLA) models for robotic manipulation. (1) Integrated speech as an end-to-end input modality in VLA training, enabling faster responses, improved performance, and enhanced user personalization [C.6].

- **Multi-agent Collaboration** *Jun. 2023 - Dec. 2023*
  Research on natural human-machine interaction technology for heterogeneous unmanned swarms
  - Designed a system enabling robots (drones and ground fleets) to understand human language and make decisions based on environmental perception. Language understanding is achieved via large language models for task decomposition and code generation. Environmental perception relies on RGB images captured by drones, which are processed by vision-language models to generate heatmaps that guide fleet actions [J.1].

Honors and Awards

- **National Scholarship** *Dec. 2024*
- **National Third Prize in the Phoenix Intelligent Technology Innovation and Application Competition** *Jun. 2021*
- **Grade A Comprehensive Scholarship in Chongqing University** *Dec. 2020*
- **Outstanding Individual Youth Volunteer of Chongqing University** *May. 2020*
- **National Scholarship** *Dec. 2019*
- **Outstanding Student of Chongqing University** *Dec. 2019*

Skills

- **Programming Languages:** Python, Pytorch, C++
- **Languages:** Chinese, English

Publications

C=Conference, J=Journal, S=In Submission

I. Generalization in Vision Language Models

[C.1] **Shuanghao Bai***, Yuedi Zhang*, Wanqi Zhou, Zhirong Luan, Badong Chen. Soft Prompt Generation for Domain Generalization. In European Conference on Computer Vision (ECCV). 2024. [Paper] [Code]

[C.2] **Shuanghao Bai**, Min Zhang, Wanqi Zhou, Siteng Huang, Zhirong Luan, Donglin Wang, Badong Chen. Prompt-based Distribution Alignment for Unsupervised Domain Adaptation. In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). 2024. [Paper] [Code]

[C.3] **Shuanghao Bai**, Wanqi Zhou, Zhirong Luan, Donglin Wang, Badong Chen. Improving Cross-domain Few-shot Classification with Multilayer Perceptron. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2024. [Paper] [Code]

[C.4] Haoran Zhang*, **Shuanghao Bai***, Wanqi Zhou, Jingwen Fu, Badong Chen. PromptTA: Prompt-driven Text Adapter for Source-free Domain Generalization. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2025. [Paper] [Code]

[S.1] Wanqi Zhou, **Shuanghao Bai**, Qibin Zhao, Badong Chen. Revisiting the Adversarial Robustness of Vision Language Models: a Multimodal Perspective. ArXiv preprint arXiv: 2404.19287. [Paper] [Code]

[S.2] Yuedi Zhang, **Shuanghao Bai**, Wanqi Zhou, Zhirong Luan, Badong Chen. Dual-Path Stable Soft Prompt Generation for Domain Generalization. ArXiv preprint arXiv: 2505.18770. [Paper] [Code]

II. Robot Learning

[C.5] **Shuanghai Bai**, Wanqi Zhou, Pengxiang Ding, Wei Zhao, Donglin Wang, Badong Chen. Rethinking Latent Representations in Behavior Cloning: An Information Bottleneck Approach for Robot Manipulation. In International Conference on Machine Learning (ICML). 2025. [Paper] [Code] [Project]

[C.6] Wei Zhao, Pengxiang Ding, Zhang Min, Zhefei Gong, **Shuanghao Bai**, Han Zhao, Donglin Wang. VLAS: Vision-Language-Action Model with Speech Instructions for Customized Robot Manipulation. In International Conference on Learning Representations (ICLR). 2025. [Paper] [Code]

[J.1] Zhirong Luan, Yijun Lai, Rundong Huang, **Shuanghao Bai**, Yuedi Zhang, Haoran Zhang, Qian Wang. Enhancing Robot Task Planning and Execution through Multi-Layer Large Language Models. In Sensors. 2024. [Paper]

[S.3] Can Cui, Pengxiang Ding, Wenxuan Song, **Shuanghao Bai**, Xinyang Tong, Zirui Ge, Runze Suo, Wanqi Zhou, Yang Liu, Bofang Jia, Han Zhao, Siteng Huang, Donglin Wang. Openhelix: A short survey, empirical analysis, and open-source dual-system vla model for robotic manipulation. ArXiv preprint arXiv: 2505.03912. [arXiv] [Code] [Project]

III. Causal Learning in Machine Learning

[C.7] Wanqi Zhou, **Shuanghao Bai**, Shujian Yu, Qibin Zhao, Badong Chen. Jacobian Regularizer-based Neural Granger Causality. In International Conference on Machine Learning (ICML). 2024. [Paper] [Code]

[J.2] Wanqi Zhou, **Shuanghao Bai**, Yicong He, Badong Chen. An Information-Theoretic Approach for Heterogeneous Differentiable Causal Discovery. In Neural Networks. 2025. [Paper] [Code]

Academic Service

- **Conference Reviewer:** ICIRA
- **Journal Reviewer:** TIP, TCSVT, KBS, NN, Neucom

About Me

As a third-year direct Ph.D. candidate at Xi'an Jiaotong University, I'm deeply fascinated by computer vision, with a particular focus on generalization in computer vision and its applications in robotics. The more I learn, the more I realize how much there is to explore in these fields!

I am actively seeking academic and industrial exchange opportunities for Fall 2025, specifically focusing on joint Ph.D. programs and internship projects. My hope is to find a research team where I can roll up my sleeves, dive into some cutting-edge projects, and both contribute my skills and learn new ones. I'm eager to experience a different academic environment and see how it shapes my perspective on research.