Ji Xie

sanaka@berkeley.edu• horizonwind2004.github.io • Google Scholar

EDUCATION

Zhejiang University, Chu Kochen Honors College *09/2022-06/2026(Expected)*

Bachelor of Engineering in Computer Science and Technology GPA: **4.20/4.3, 93.70/100** | Rank**: 2/148**

RESEARCH INTEREST

My long-term goal is to build a unified, controllable, and **powerful multimodal model** and apply its **strong generative priors** to the construction of **world models** and **embodied systems.**

SELECTED RESEARCH EXPERIENCE

Unified Multimodal Model 03/2025-Now **UC Berkeley**

- Pioneered a novel post-training paradigm of *Unified Multimodal Model* (RecA). **First to propose** using vision embeddings as *"dense prompts"* to enforce semantic consistency without text annotations.
- Our **1.5B model achieves SOTA** results, boosting performance with merely 27 A100 GPU-hours.
- Investigating how generative objectives can serve as supervisory signals to enhance the understanding and representation capabilities of the *Unified Multimodal Model*.
- Advisers: Trevor Darrell, XuDong Wang.

Controllable Image Generation 01/2024-03/2025 **Zhejiang University**

- First to treat instructional image editing as an in-context learning task (ICEdit), achieving **SOTA results** with **only 0.1% data** of existing baselines. Surpassed *GPT-4o-Image* and *SeedEdit* in ID consistency.
- Devised a novel depth-driven generation framework (3DIS) that serves as **a universal plug-and-play module** across diverse diffusion backbones. Pioneered the use of depth maps for precise spatial grounding, setting a new SOTA for layout fidelity and visual quality.
- Advisers: Dewei Zhou, Zechuan Zhang, Yi Yang.

SELECTED PUBLICATIONS

- **Ji Xie**, Trevor Darrell, Luke Zettlemoyer, XuDong Wang, "Reconstruction Alignment Improves Unified Multimodal Models," under review on *International Conference on Learning Representation* (*ICLR*), 2026 (8864, **top 2.5%**). **Code.**
- Zechuan Zhang, **Ji Xie**, Yu Lu, Zongxin Yang, Yi Yang, "Enabling Instructional Image Editing with In-Context Generation in Large Scale Diffusion Transformer," *Conference on Neural Information Processing Systems (NeurIPS)*, 2025. **Code (2,000 Stars!).**
- Yanhao Jia, **Ji Xie**, S Jivaganesh, Hao Li, Xu Wu, Mengmi Zhang. "Seeing Sound, Hearing Sight: Uncovering Modality Bias and Conflict of AI Models in Sound Localization," *Conference on Neural Information Processing Systems (NeurIPS)*, 2025, **Spotlight**. **Code.**
- Dewei Zhou*, **Ji Xie***, Zongxin Yang, Yi Yang, "3DIS: Depth-Driven Decoupled Image Synthesis for Universal Multi-Instance Generation," *International Conference on Learning Representation* (*ICLR*), 2025, **Spotlight**. **Code.**

SELECTED COMPETITION & AWARDS

Outstanding **Peer Tutor Award**, Zhejiang University, 2025 *06/2025*

SenseTime Scholarship (**top 30 recipients annually in China**) *06/2025*

Zhejiang Provincial Government Scholarship *12/2024*
International Collegiate Programming Contest (ICPC), Shenyang Site, 2022, **Gold Medal** *10/2022*
China Collegiate Programming Contest (CCPC), Guangzhou Site, 2022, **Gold Medal** *10/2022*
Zhejiang Provincial Collegiate Programming Contest, 2024, **Gold Medal** *04/2024*
Zhejiang Provincial Higher Mathematics Competition, 2024, **First Prize** *06/2024*

VOLUNTEER ACTIVITIES

Chu Kochen College Academic Guidance Center 09/2023-06/2024 **Zhejiang University**

- Organized peer tutoring sessions for Calculus and Mathematical Analysis; designed original lecture slides to help students understand the intuition behind complex formulas.
- Co-authored the open-source lecture LALU, receiving over **600 stars** on GitHub.
- Produced and uploaded educational videos to Bilibili, accumulating thousands of views.

Summer Volunteer at Shanghai AI Lab 08/2023-09/2023 **Shanghai**

- Taught high school students the basics of Python programming and Large Language Models (LLMs).
- Guided students to build their own simple AI agents using OpenAI APIs, helping them gain hands-on experience.

PROFESSIONAL SKILLS

Large-scale Distributed Training: Extensive experience in training Large Language Models / Generative Models on high-performance clusters (256 NVIDIA H20 GPUs).

Core Stack: Python, C/C++, PyTorch, NumPy, Matplotlib, Linux.

Codeforces: **2478** (id: Epyset).

Good at Mathematics and Physics. Linear Algebra (100), Mathematical Analysis I (96), Mathematical Analysis II (96), Probability Theory and Mathematical. Statistics (99), General Physics lI (100).