Jinrui Yang

jinruiyang.ray@gmail.com · Google Scholar · Semantic Scholar

Recent research interests

My current research interests lie in the field of generative models, multimodal large language models, and large language models.

Education

University of California, Santa Cruz, CA, U.S. 2023.08 – present

Ph.D. student in Computer Science. Advisor: Cihang Xie and Yuyin Zhou

Sun Yat-sen University, Guangzhou, China 2019.09 – 2021.06

M.E. in Computer Technology.

Sichuan University, Chengdu, China 2015.09 – 2019.06

B.E. in Software Engineering

Work Experience

ByteDance Seed. San Jose, USA 2025.10 – Present

Research Intern

- **Research Topic: Native Multimodal Pre-training.**
- Investigating a unified generative paradigm to bridge symbolic text, visual-text, and visual perception, enabling the seamless and interleaved generation of symbolic text, visual-text, and natural images, and unlocking new forms of multimodal knowledge acquisition from visually rich data such as PDFs.

Adobe Research. San Jose, USA 2025.06 – 2025.10

Research Intern

- **Research Topic: Spatial Reasoning in MLLMs.**
- Proposed a novel dataset to enhance the occlusion reasoning capabilities of MLLMs via RL.

Adobe Research. San Jose, USA 2024.06 – 2025.06

Research Intern

- **Generative Image Layer Decomposition with Visual Effects:** Designed **LayerDecomp**, a layered image decomposition method that preserves transparent visual effects and enables fine-grained editing, powered by a scalable synthetic dataset pipeline, resulting in a paper accepted at CVPR 2025.
- **Controllable Layered Image Generation for Real-World Editing:** Developing a layered image generation framework for controllable RGBA synthesis with realistic visual effects, along with a supporting dataset.

Tencent YouTu Lab. Shanghai, China 2021.07 – 2023.08

Research Scientist, Full-time

- Built robust vision perception models for different business scenarios.
- Applied MLLMs to downstream visual tasks in the real-world.
- Built the comprehensive MLLM evaluation benchmark MME, which has been widely adopted by mainstream multimodality models (e.g., LLaVA, Qwen-VL, InternVL).

Tencent YouTu Lab. Shanghai, China 2020.05 – 2020.10

Research Intern

Conducted research on person re-identification, resulting in a paper accepted at ICCV 2021.

Publications

1. **Jinrui Yang**, Qing Liu, Yijun Li, Soo Ye Kim, Daniil Pakhomov, Mengwei Ren, Jianming Zhang, Zhe Lin, Cihang Xie, Yuyin Zhou. **Generative Image Layer Decomposition with Visual Effects**. **CVPR 2025**. Project page.

2. Chaoyou Fu, Peixian Chen, Yunhang Shen, Yulei Qin, Mengdan Zhang, Xu Lin, **Jinrui Yang**, Xiawu Zheng, Ke Li, Xing Sun, Yunsheng Wu, Rongrong Ji. **MME: A Comprehensive Evaluation Benchmark for Multimodal Large Language Models**. **NeurIPS 2025 (Highlight)**. Paper.
3. **Jinrui Yang**, Xianhang Li, Druv Pai, Yuyin Zhou, Yi Ma, Yaodong Yu, Cihang Xie. **Scaling White-Box Transformers for Vision**. **NeurIPS 2024**. Project page.
4. **Jinrui Yang**, Jiawei Zhang, Fufu Yu, Xinyang Jiang, Mengdan Zhang, Xing Sun, Yingcong Chen, Wei-Shi Zheng. **Learning to Know Where to See: A Visibility-Aware Approach for Occluded Person Re-identification**. **ICCV 2021**. Paper.
5. **Jinrui Yang**, Wei-Shi Zheng, Qize Yang, Yingcong Chen, Qi Tian. **Spatial-Temporal Graph Convolutional Network for Video-Based Person Re-identification**. **CVPR 2020**. Paper.
6. Yuqiao Xian, **Jinrui Yang**, Fufu Yu, Jun Zhang, Xing Sun. **Graph-Based Self-Learning for Robust Person Re-identification**. **WACV 2023**. Paper.

PREPRINTS

1. **Jinrui Yang**, Qing Liu, Yijun Li, Mengwei Ren, Letian Zhang, Zhe Lin, Cihang Xie, Yuyin Zhou. **Controllable Layered Image Generation for Real-World Editing**. (**Under review**) Project page.
2. **Jinrui Yang**, Zonglin Di, Ohi Dibua, Qing Liu, Seun Adekunle, Daniil Pakhomov, Darshan Ganesh Prasad, Cihang Xie, Yuyin Zhou. **Improving the Capability of Visual Language Models with Occlusion Reasoning**. (**Under review**)
3. Jiaming Zhou, Junwei Liang, Kun-Yu Lin, **Jinrui Yang**, Wei-Shi Zheng. **ActionHub: A Large-scale Action Video Description Dataset for Zero-shot Action Recognition**. **arXiv 2024**. Paper

ACADEMIC ACTIVITIES

Conference Reviewer: AAAI2026, NeurIPS2025, CVPR2025, ICML2024, CVPR2024, WACV2023, WACV2025.
Journal Reviewer: TIP, TCSVT, TMM