

Shuo Cao

PhD Candidate, USTC

L1, International Media Port, Longwen Road

Xuhui, Shanghai, China

☎ (+86) 15655601206

✉ caoshuo@pjlab.org.cn

🏠 [Homepage](#) / 🎓 [Google Scholar](#) / 🐙 [GitHub](#)

Short Biography

I am currently a third-year joint PhD student at the University of Science and Technology of China and Shanghai AI Laboratory, supervised by Prof. Yu Qiao. I received my bachelor's degree from Tongji University in 2023. Previously, I worked on low-level vision with Prof. Chao Dong. My current research interests focus on multimodal image understanding and unified models for image generation and understanding. I am now working closely with Dr. Yihao Liu and Prof. Yu Qiao.

Education

- 09/2023– Present **PhD Student**, *University of Science and Technology of China*, Hefei, China, Supervised by Prof. Yu Qiao.
- 09/2019– 06/2023 **B.E. in Automation**, *Tongji University*, Shanghai, China, School of Electronics and Information Engineering; GPA 4.72/5.0, Ranking 2/78.

Research Experience

- 09/2024– Present **Research Intern**, *Shanghai AI Laboratory*, Shanghai, China.
- Member of OpenGVLab, supervised by Dr. Yihao Liu and Prof. Yu Qiao. Research in multi-modal models, especially in low-level vision, multimodal image understanding, reward models & evaluation, and unified models for image generation and understanding (ongoing).

Research Highlights

- ★ **Low-Level Vision:** Representative works include **GRIDS (ECCV 2024)** for grouped multiple-degradation restoration, **DualX-VSR** for compensation-free real-world video super-resolution, and collaborative works such as **DiffVSR (ICCV 2025)**, **LinearSR (ICLR 2026)**.
- ★ **Multimodal Image Understanding:** Representative works include **UniPercept (ICML 2026 Spotlight)** for unified perceptual-level understanding across aesthetics, quality, structure, and texture; **ArtiMuse (CVPR 2026)** for fine-grained aesthetic assessment with expert-level understanding; and **StableI2I (ICML 2026)** for fidelity-oriented image-to-image transition evaluation. These works have been widely validated as practical evaluators and effective reward models.
- ★ **Unified Generation-Understanding Foundation Model (ongoing):** I am currently working on full-stack model development across model design, data construction, training recipe, and large-scale pretraining. I collaborated with Huawei engineers to implement the full pretraining pipeline and accuracy alignment on **1,024 Ascend 910B NPUs**.

Selected Publications (First & Co-first Author)

- ★ **Shuo Cao***, Jiayang Li*, Xiaohui Li, Yuandong Pu, . . . , Bin Fu, Yu Qiao, Yihao Liu. UniPercept: Towards Unified Perceptual-Level Image Understanding across Aesthetics, Quality, Structure, and Texture. *International Conference on Machine Learning (ICML)*, 2026. **Spotlight**.
- ★ Jiayang Li*, **Shuo Cao***, Xiaohui Li, Zhenxing Zhang, Kaiwen Zhu, Yuran Duan, Yu Qiao, Jian Zhang, Yihao Liu. StableI2I: Spotting Unintended Changes in Image-to-Image Transition. *International Conference on Machine Learning (ICML)*, 2026.

- ★ **Shuo Cao**, Nan Ma, Jiayang Li, Xiaohui Li, . . . , Yu Qiao, Dajun Yao, Yihao Liu. *ArtiMuse: Fine-Grained Image Aesthetics Assessment with Joint Scoring and Expert-Level Understanding*. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.
- ★ **Shuo Cao***, Yihao Liu*, Wenlong Zhang, Yu Qiao, Chao Dong. *GRIDS: Grouped Multiple-Degradation Restoration with Image Degradation Similarity*. *European Conference on Computer Vision (ECCV)*, 2024.
- ★ Yuanting Gao*, **Shuo Cao***, Xiaohui Li, Yuandong Pu, Yihao Liu, Kai Zhang. *Toward Generalizable Deblurring: Leveraging Massive Blur Priors with Linear Attention for Real-World Scenarios*. *Under Review*, 2026.
- ★ **Shuo Cao***, Yihao Liu*, Xiaohui Li, Yuandong Pu, Yihao Liu, Kai Zhang. *DualX-VSR: Dual Axial Spatial \times Temporal Transformer for Real-World Video Super-Resolution without Motion Compensation*. *Under Review*, 2025.

Selected Publications (Collaborative)

- ★ Xiaohui Li, Shaobin Zhuang, **Shuo Cao**, Yang Yang, . . . , Bin Fu, Yihao Liu. *LinearSR: Unlocking Linear Attention for Stable and Efficient Image Super-Resolution*. *International Conference on Learning Representations (ICLR)*, 2026.
- ★ Yuandong Pu, Le Zhuo, Songhao Han, Jinbo Xing, Kaiwen Zhu, **Shuo Cao**, . . . , Yu Qiao, Wenlong Zhang, Xi Chen, Yihao Liu. *PICA-Bench: How Far Are We from Physically Realistic Image Editing?* *International Conference on Learning Representations (ICLR)*, 2026.
- ★ Yi Xin, Siqi Luo, Tianxiang Xu, Qi Qin, . . . , **Shuo Cao**, . . . , Yihao Liu, Yuewen Cao, Xiaohong Liu. *dMLLM-TTS: Self-Verified and Efficient Test-Time Scaling for Diffusion Multi-Modal Large Language Models*. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.
- ★ Xiaohui Li, Yihao Liu, **Shuo Cao**, Ziyang Chen, . . . , Yi Wang, Yu Qiao. *DiffVSR: Revealing an Effective Recipe for Taming Robust Video Super-Resolution against Complex Degradations*. *International Conference on Computer Vision (ICCV)*, 2025.
- ★ Kaiwen Zhu, Quansheng Zeng, Yuandong Pu, **Shuo Cao**, . . . , Jinjin Gu, Yihao Liu. *Accelerating Masked Image Generation by Learning Latent Controlled Dynamics*. *Under Review*, 2026.
- ★ Xiangyu Chen, Kaiwen Zhu, Yuandong Pu, **Shuo Cao**, . . . , Yu Qiao, Jiantao Zhou, Chao Dong. *Exploring Scalable Unified Modeling for General Low-Level Vision*. *Under Review*, 2025.

Projects

- ★ **ArtiMuse & UniPercept Series**: Released code, checkpoints, benchmarks, datasets, and online/project resources for perceptual-level image understanding. The series has received **320+ GitHub stars**; UniPercept reached **60K+ downloads** within six months after release.
- ★ **ArtiMuse Online Demo**: Collaborated with China Academy of Art to deploy the [Online Demo](#) of ArtiMuse, serving **500K+ model/API calls** within four months after release.

Talks

- 03/2026 Towards Unified Perceptual-level Image Understanding: From ArtiMuse to UniPercept, Apple.
- 11/2025 MLLMs for Expert-Level Image Aesthetics Assessment, Huawei.
- 08/2025 Invited talk on *ArtiMuse: Fine-Grained Image Aesthetics Assessment with Joint Scoring and Expert-Level Understanding* at the forum “Evolving with AI: The Iteration and Resilience of Artistic Creativity” during the **World Artificial Intelligence Conference (WAIC) 2025**.

Awards & Honors

First-Class Undergraduate Scholarship (TOP 5%).

National First Prize, China Undergraduate Mathematical Contest in Modeling (Top 1%).