Xuhan Sheng

llstela.sxh@gmail.com | llstela.github.io | GitHub: llstela | Google Scholar

Ph.D. student in Artificial Intelligence, Institute of Science Tokyo.

Research background in diffusion models, memory-efficient training and inference, vision-language models, and video data pipelines. Co-first author of an ECCV 2024 Oral paper.

Research interests: video generation, AI agents, and large language models.

Seeking: Long-term research internships; available to start as soon as arrangements permit. Full-time availability subject to local work authorization and company policy.

Education

Institute of Science Tokyo, Department of Computer Science Oct. 2026 – Present

Ph.D. student in Artificial Intelligence | Advisor: Prof. Rio Yokota
Earliest graduation: Sept. 2029; option to graduate in Apr. 2030.

Peking University, Computer Application Technology Sept. 2023 – Jun. 2026

Master's degree | GPA: 3.87/4.00 | Advisor: Prof. Jian Zhang

Dalian University of Technology, Artificial Intelligence Sept. 2019 – Jun. 2023

Bachelor's degree | GPA: 4.17/5.00 | Advisor: Prof. Xu Jia

Research and Industry Experience

Xiaomi Auto, Shanghai Team — World Model Research Intern Aug. 2026 – Present

- Contributing to **Xiaomi Auto World Model 2.0** (forthcoming), with primary responsibility for **video super-resolution**.

Tencent IEG, Game AI Engine Department — Algorithm Intern Mar. 2026 – Jun. 2026

- Developed Qwen-Image-based super-resolution for game visuals and production assets.
- Built data-pipeline operators for video generation, including streaming shot segmentation that converts long videos into single-shot clips; work was integrated into the production pipeline.

OPPO Research Institute — Imaging Algorithm Engineer Intern May 2025 – Sept. 2025

- Researched vision-language models for locating and annotating perceptual artifacts in AI-generated images to support image enhancement. AI Talent Program; advisor: Prof. Lei Zhang.

Selected Research and Publications

* Equal contribution. Full publication list: llstela.github.io.

OmniSSR: Zero-shot Omnidirectional Image Super-Resolution using Stable Diffusion Model
Runyi Li*, **Xuhan Sheng***, Weiqi Li, Jian Zhang. **ECCV 2024, Oral**.

- Contributed to method design, implementation, and evaluation of zero-shot panoramic super-resolution using diffusion priors, tangent-view information exchange, and gradient decomposition.

PanoForcing: Enforcing Memory-Scalable Full-Context for Blind Panoramic Super-Resolution
Xuhan Sheng (first author), et al. **Under review, ICLR 2027**.

- Led method design, implementation, core experiments, and manuscript writing. Combined selective autograd subgraph routing with Fourier pseudoinverse guidance for memory-scalable, full-context one-step panoramic super-resolution.

RealOSR: Latent Guidance Boosts Diffusion-based Real-world Omnidirectional Image Super-Resolutions
Xuhan Sheng*, Runyi Li*, Bin Chen, Weiqi Li, Xu Jiang, Jian Zhang. **Preprint**, arXiv:2412.09646.

- Major contributor to implementation, experiments, and manuscript writing for one-step diffusion-based panoramic restoration with latent-space guidance.

OmniDrag: Enabling Motion Control for Omnidirectional Image-to-Video Generation
Weiqi Li, Shijie Zhao, Chong Mou, **Xuhan Sheng**, et al. **IJCV 2025**.

- Contributed to the Move360 large-motion dataset for controllable panoramic image-to-video generation.

Label-guided Facial Retouching Reversion
Guanhua Zhao*, Yu Gu*, **Xuhan Sheng**, Yujie Hu, Jian Zhang. **ICME 2025, Oral**.

- Developed hierarchical adaptive instance normalization to mitigate color drift in facial retouching reversion.

Skills and Recognition

Programming: Python, PyTorch. **English:** IELTS Academic 7.5/9.0.
Awards: NTIRE 2023 omnidirectional SR challenge winning team; PKU Academic Excellence Award (2024); DUT University-level Outstanding Graduate (2023).