

XUHAN SHENG (盛栩涵)

MB.(Wechat): (+86) 13-732463456
Email: shengxuhan@stu.pku.edu.cn

Google Scholar: user=7Sfji6YAAAAJ
GitHub: llstela.github.io

EDUCATION

Institute of Science Tokyo, Department of Computer Science Starting from Oct. 2026

- Ranked **18th globally** and **1st in Japan** in Computer Science (QS Subject Rankings 2026).
- Incoming Ph.D. Student | Computer Science | Advisor: Rio Yokota

Peking University, School of Electronic and Computer Engineering 2023 – 2026

- Master's Degree | Computer Application Technology | GPA: 3.87/4.00 | Advisor: Jian Zhang (VILLA)

Dalian University of Technology, School of Artificial Intelligence 2019 – 2023

- Bachelor's Degree | Artificial Intelligence | GPA: 4.17/5.00 | Advisor: Xu Jia (IIAU-LAB)

INTERNSHIPS

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

- AI Talent Program: Class of 2026 Dream-Seeking Internship
- Vision-Language Models for Assisting AI-based Image Enhancement.
- Advisor: Lei Zhang (IEEE Fellow, Chair Professor of Hong Kong Polytechnic University)

Tencent IEG, Game AI Engine Department Intern Mar. 2026 – Current

- Image Super-Resolution based on Qwen-Image for game content enhancement.
- Developed and optimized AI-based image restoration techniques for production-level visual assets.
- Related techniques have been deployed and adopted in real business scenarios.

RESEARCH BACKGROUND

(1) **Diffusion-based Super-Resolution for Panoramic Images:** Panoramic images, pivotal in VR/AR applications, necessitate ultra-high resolution (e.g., 4K×8K), yet real-world degradations make reconstruction highly ill posed. Diffusion models provide strong image priors, enabling high fidelity restoration.

(2) **Vision-Language Models for Assisting AI-based Image Enhancement:** AI generated images often contain perceptually significant artifacts. A vision-language model is trained to precisely localize artifact regions with interpretable annotations, thereby facilitating more accurate super-resolution and broader enhancement tasks.

FUTURE RESEARCH INTEREST

(1) **Efficient Vision Encoder:** with dynamic vision token compression/pruning for real-time control VLA models.

(2) **Dual VLA systems:** with high-level models for reasoning and low-level models for execution, tractable for long-horizon tasks and cross-embodiment.

PUBLICATIONS

(* denotes co-first author)

- **First Author Papers** (ECCV'24 ORAL; UNDER REVIEW)

1. **RealOSR: Latent Unfolding Boosts Diffusion-based Real-world Omnidirectional Image Super-Resolution**

Xuhan Sheng*, Runyi Li*, Bin Chen, Weiqi Li, Xu Jiang, Jian Zhang.

Under Review.

– **Overview:** One-step diffusion-based omnidirectional image super-resolution guided by latent space unfolding, with lightweight domain alignment and degradation-aware modules for real-world restoration.

– **Contributions:** Major contributor to codebase, experiments, and manuscript writing.

2. **OmniSSR: Zero-shot Omnidirectional Image Super-Resolution using Stable Diffusion Model**

Runyi Li*, Xuhan Sheng*, Weiqi Li, Jian Zhang.

ECCV 2024 (Oral, Acceptance Rate: 2.3%) (Top-tier international conference in computer vision).

- **Overview:** First zero-shot omnidirectional image super-resolution method leveraging Stable Diffusion priors, integrating Octadecaplex Tangent Information Interaction and Gradient Decomposition without training.
 - **Contributions:** Code implementation, method design, and experimental evaluation.
- **Other Contributions** (IJCV'25; ICME'25 ORAL; PRCV'25; CVPRW'23 NTIRE CHAMPION)
 1. **OmniDrag: Enabling Motion Control for Omnidirectional Image-to-Video Generation**
 Weiqi Li, Shijie Zhao, Chong Mou, **Xuhan Sheng**, et al.
IJCV 2025 (Top-tier journal in computer vision).
 - **Overview:** First diffusion-based framework enabling precise drag-style motion control for omnidirectional image-to-video generation.
 - **Contributions:** A large-motion dataset construction (Move360).
 2. **Bridging the Point to Boundary Gap for Point-supervised Temporal Action Localization with Single-stage Inference**
 Junshi Yang, Shenglan Liu, **Xuhan Sheng**, et al.
PRCV 2025 (The largest and most comprehensive technical conference in China, focusing on pattern recognition).
 - **Overview:** A hierarchical framework converting sparse point annotations into boundary-accurate pseudo labels via Gaussian-prior boundary enhancement.
 - **Contributions:** Assisted in paper writing and submission.
 3. **Label-guided Facial Retouching Reversion**
 Guanhua Zhao*, Yu Gu*, **Xuhan Sheng**, Yujie Hu, Jian Zhang.
ICME 2025 (Oral) (Top-tier international conference in multimedia).
 - **Overview:** A diffusion-based method for reversing facial retouching effects such as skin smoothing, eye enlargement, and face slimming.
 - **Contributions:** Developed hierarchical adaptive instance normalization to mitigate color drift, achieving >20 FID improvement and +1.4 dB PSNR.
 4. **OPDN: Omnidirectional Position-Aware Deformable Network for Omnidirectional Image Super-Resolution**
 Xiaopeng Sun*, Weiqi Li*, Zhenyu Zhang, Qiufang Ma, **Xuhan Sheng**, et al.
CVPRW 2023 (NTIRE 360° Panoramic SR Challenge Champion).
 - **Overview:** Champion solution for the NTIRE 2023 Challenge on 360° panoramic image super-resolution.
 - **Contributions:** Proposed a spatial-frequency fusion module and collected fisheye camera data to support model training.

PRACTICAL SKILLS

- Programming: Python (with PyTorch)

ENGLISH PROFICIENCY

- DET: 120 / 160
- IELTS Academic: 7.5 / 9.0

HONORS AND AWARDS

- Peking University **Academic Excellence Award** 2024
- Dalian University of Technology **University-level Outstanding Graduate** 2023
- Dalian University of Technology **First-class Academic Excellence Scholarship** 2022
- Dalian University of Technology **Second-class Academic Excellence Scholarship** 2021
- Dalian University of Technology **First-class Academic Excellence Scholarship** 2020