Xinjie Li

Department of Computer Science and Engineering
Pennsylvania State University–University Park
☎ +1 (582) 203-8166 • ✉ xql5497@psu.edu

Research Interests

Recent Focus: Video MLLM (Omni) and Diffusion; Reinforcement Learning
Earlier Background: (Human) Video/Audio/Image/Motion Generation, Video Understanding, Image Restoration, AI4Science, Transformer, Long-tailed Modeling, Metric Learning, Segmentation, and Detection.

Education

Pennsylvania State University–University Park 2022-2026
Ph.D., Computer Science (University Graduate Fellowship Recipient) *State College, PA*
Advisor: Prof. Abhinav Verma

University of Science and Technology Beijing 2019-2022
M.S., Computer Science (GPA: 3.4/4.0) *Beijing, China*
Advisor: Prof. Xu-Cheng Yin

University of Science and Technology Beijing 2015-2019
B.S., Automation (Honors College) (Ranking: 5/53) (GPA: 3.7/4.0) *Beijing, China*

Publications

Video Understanding

[1]. CI w/o TN: Context Injection without Task Name for Procedure Planning.
Xinjie Li.
Arxiv Preprint.

[2]. Repetitive Action Counting with Motion Feature Learning.
Xinjie Li, Huijuan Xu.
IEEE/CVF Winter Conference on Applications of Computer Vision (**WACV**), 2024.

[3]. MEID: Mixture-of-Experts with Internal Distillation for Long-Tailed Video Recognition.
Xinjie Li, Huijuan Xu.
37th AAAI Conference on Artificial Intelligence (**AAAI**), 2023. (**Oral**)

Image/Video Generation

[1]. Co-Speech Gesture Video Generation Implicit Motion-Audio Entanglement.
Xinjie Li, Ziyi Chen, Xinlu Yu, Iek-Heng Chu, Peng Chang, Jing Xiao
IEEE/CVF Conference on Computer Vision and Pattern Recognition **CVPR**, 2025.

[2]. Are Conditional Latent Diffusion Models Effective for Image Restoration?
Yunchen Yuan, Junyuan Xiao, **Xinjie Li**†
Arxiv Preprint.

[3]. DiffBody: Human Body Restoration by Imagining with Generative Diffusion Prior.
Yiming Zhang, Zhe Wang, **Xinjie Li**, Yunchen Yuan, Chengsong Zhang, Xiao Sun, Zhihang Zhong, Jian Wang
Arxiv Preprint. Submitted to ICCP 2025

[4]. SD-GAN: Semantic Decomposition for Face Image Synthesis with Discrete Attribute.
Kangneng Zhou, Xiaobin Zhu, Daiheng Zhou, Kai Lee, **Xinjie Li**, Xu-Cheng Yin.
30th ACM International Conference on Multimedia (**ACM MM**), 2022.

AI4Science

[1]. RetroMoE: A Mixture-of-Experts Latent Translation Framework for Single-step Retrosynthesis.
Xinjie Li, Abhinav Verma.
34th International Joint Conference on Artificial Intelligence (**IJCAI**), 2025

[2]. Sub-Clustering for Class Distance Recalculation in Long-Tailed Drug Discovery.
Yujia Su*, **Xinjie Li***†, Lionel Z. Wang.
Arxiv Preprint.

[3]. EllipBench: A Large-scale Benchmark for Machine-learning based Ellipsometry Modeling.
Yiming Ma*, **Xinjie Li*†**, Xin Sun, Zhiyong Wang, Lionel Z. Wang.
Arxiv Preprint. Submitted to NeurIPS 2025 Datasets & Benchmarks Track

Others

[1]. Stereo CenterNet based 3D Object Detection for Autonomous Driving.
Yuguang Shi, Yu Guo, Zhenqiang Mi, **Xinjie Li**.
Neurocomputing, 2022.

[2]. DANet: Dynamic Attention to Spoof Patterns for Face Anti-Spoofing.
Chun-Yu Sun, Song-Lu Chen, **Xinjie Li**, Feng Chen, Xu-Cheng Yin.
26th International Conference on Pattern Recognition (**ICPR**), 2022.

[3]. Semantic Bilinear Pooling for Fine-Grained Recognition.
Xinjie Li, Chun Yang, Song-Lu Chen, Chao Zhu, Xu-Cheng Yin.
24th International Conference on Pattern Recognition (**ICPR**), 2020.

Research Internships

Amazon Prime Video. (US) — May 2025 -
Video Understanding and Generation — *Seattle, WA*
Advisor: Dr. Zhimin Chen, Dr. Florian Schiffers, Dr. Rui Zhao

PAII Inc. (US) — May 2024 - March 2025
Co-Speech Gesture Video Generation — *Palo Alto, CA*
Advisor: Dr. Ziyi Chen, Dr. Peng Chang

Center for AI Safety. (US) — Feb 2024 - Apr 2024
EBM enhanced Language Model for Text Generation — *Remote*
Advisor: Dr. Xuwang Yin

Snap Research. (US) — Aug 2023 - March 2024
ControlNet based Human Image Restoration — *Remote*
Advisor: Dr. Jian Wang

Computational Biology Lab. (CMU) (US) — May 2023 - Jan 2024
Diffusion for Cryo-ET Image Synthesis and Missing-Wedge Correction — *Remote*
Advisor: Dr. Xiangrui Zeng, Prof. Min Xu

Meituan Inc. (CN) — May 2021 - Aug 2021
Vision Transformer and Video Semantic Segmentation — *Beijing, China*
Advisor: Xiangxiang Chu

Remark AI. (UK) — Feb 2021 - Apr 2021
Retail Product Automatic Checkout — *Remote*
Advisor: Dr. Xiaoyun Yang

USTB and EEasy Tech Joint Lab. (CN) — Sep 2017 - Jan 2021
Fine-grained and Long-tailed Image Classification, Vehicle Re-identification — *Beijing, China*
Advisor: Prof. Xu-Cheng Yin

Projects

Next Token Prediction in Computer Vision — May 2023 - July 2023
Finetune Stable Diffusion Inpainting model with Dreambooth and Lora on SA1B dataset.

Competitions

◇ ICCV 2021 Video Scene Parsing in the Wild Challenge (Top 10)

◇ CVPR 2021 FGVC 8 Plant Pathology Challenge (Top 10%)

◇ CVPR 2020 FGVC 7 Semi-Supervised Recognition Challenge (Top 15)

Professional Activities

Conference Reviewer/Program Committee Member

IJCAI 2025; ICCV 2025; ICML 2025; ICLR 2025; NeurIPS 2024; CVPR 2023, 2024, 2025; ECCV 2024; AAAI 2023, 2024, 2025; ACM MM 2022, 2023, 2024; WACV 2023