Xinjie Li

+1 (582) 203-8166 • abcdvzz@hotmail.com
scholar.google.com/citations?user=sh9uq8EAAAAJ

Current Appointment

Amazon Prime Video — **Seattle, WA**
Applied Scientist — *Jan. 2026 – Present*

Research Interests

Recent Focus: MLLM Post-Training; Agent

Primary Expertise: Unified Multimodal Video Understanding (Omni MLLM) and Generation (Diffusion); Reinforcement Learning

Background: Generative Models, MoE, (Human) Video/Audio/Image/Motion Understanding and Generation, Image Restoration, AI4Science, Long-tailed Modeling, Metric Learning

Education

Pennsylvania State University–University Park — **State College, PA**
Ph.D. in Computer Science — *2022–2026*
University Graduate Fellowship Recipient
Topic: Human Video Understanding and Generation
Advisor: Prof. Abhinav Verma

University of Science and Technology Beijing — **Beijing, China**
M.S. in Computer Science and Technology — *2019–2022*
Topic: Fine-grained and Long-tailed Image Classification, Vehicle Re-identification
Advisor: Prof. Xu-Cheng Yin

University of Science and Technology Beijing — **Beijing, China**
B.S. in Automation — *2015–2019*

Internships

Amazon Prime Video — **Seattle, WA**
Applied Scientist Intern — *May 2025 – Present*
Project: Unified Video Und and Gen Model – Large Scale Pretraining, SFT, RL
Output: One paper accepted at CVPRW 2026
Advisors: Dr. Zhimin Chen, Dr. Zhenyu Liao, Dr. Florian Schiffers, Dr. Rui Zhao

PAII Inc. **Palo Alto, CA**
Research Scientist Intern *May 2024 – Mar 2025*
Project: Co-Speech Gesture Video Generation
Output: One paper accepted at CVPR 2025, one under review at ECCV 2026
Advisors: Dr. Ziyi Chen, Dr. Peng Chang

Meituan Inc. **Beijing, China**
MLE Intern *May 2021 – Aug 2021*
Project: Vision Transformer and Video Semantic Segmentation
Advisor: Xiangxiang Chu

Preprints

(* indicates equal contribution; † indicates project lead.)

[1]. Weakly Supervised Motion Learning for Co-speech Gesture Video Generation.
Xinjie Li, Ziyi Chen, Xinlu Yu, Peng Chang, Jing Xiao, Abhinav Verma.
Submitted to **ECCV** 2026.

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

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

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

[5]. Modeling Inverse Ellipsometry Problem via Flow Matching with a Large-Scale Dataset.
Yiming Ma*, Jianzhi Teng*, **Xinjie Li**†*, Xin Sun, Zhiyong Wang, Yuzhou Song, Lionel Z. Wang, Bin Chen.
Arxiv Preprint.

Peer-Reviewed Conference Papers

[1]. What Happens Next? Next Scene Prediction with a Unified Video Model.
Xinjie Li, Zhimin Chen, Rui Zhao, Florian Schiffers, Zhenyu Liao, Vimal Bhat.
IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshop (**CVPRW**), 2026. (**Oral**)

[2]. Co-Speech Gesture Video Generation with 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.

[3]. DiffBody: Human Body Restoration by Imagining with Generative Diffusion Prior.
Yiming Zhang, Lionel Z. Wang, Sizhuo Ma, **Xinjie Li**, Jian Ren, Zhihang Zhong, Jian Wang.
IEEE International Conference on Computational Photography (**ICCP**), 2025.

[4]. 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.

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

[6]. 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**)

[7]. 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.

[8]. 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.

[9]. 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.

Peer-Reviewed Journal Articles

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

Professional Services

Conference Reviewer / Program Committee Member: CVPR 2023–2026;

ICLR 2025, 2026; NeurIPS 2024–2026; ECCV 2024, 2026; AAAI 2023–2026; BMVC 2026; ACL ARR 2026; IJCAI 2025; ICCV 2025; ICML 2025; ACM MM 2022–2024; WACV 2023

Theses

Xinjie Li. "Implicit Motion Learning for Co-speech Gesture Video Generation". Ph.D. Thesis, The Pennsylvania State University, 2026.

Patents

CN113449613B: Multi-task long-tail distribution image recognition method, 2024.
CN108082318B: Six-rod thirty-rope flexible stretching whole robot, 2023.
CN114494291A: Image segmentation method and device based on attention model, 2022.
CN108898060A: CNN model recognition method under vehicle environment, 2018.
CN207809567U: A kind of flexible tensioning entirety robot of six bars, 30 ropes, 2018.
CN206684544U: Vehicle-mounted and indoor dual-purpose high-temperature low-oxygen induction alarm device, 2017.

Teaching Experience

CMPSC/Math 455 Introduction to Numerical Analysis I, Spring 2026 & 2025;
CMPEN/EE 454 Fundamentals of Computer Vision, Fall 2025 & 2024;
CMPSC 448 Machine Learning and Algorithmic AI, Spring 2024;
CMPSC 131 Programming and Computation I: Fundamentals (Python), Fall 2023;

Last updated: May 2026