Xiang Li

📱(412) 600 2748　✉ xl6@andrew.cmu.edu　🔗lxa9867.github.io

EDUCATION

Ph.D. in Electrical and Computer Engineering, Carnegie Mellon University　　2021-2025
Supervised by Prof. Bhiksha Raj (Full professor at Language Technologies Institute, SCS)
Research Interests: Multimodal Perception / Generation
M.S. in Electrical and Computer Engineering, Carnegie Mellon University　　2021-2022
B.S. in Electrical Engineering, Huazhong University of Science and Technology　　2016-2020

RESEARCH

I am working on image/video generation with autoregressive/diffusion models and their corresponding autoencoder/tokenizer.

EXPERIENCE

Research Scientist, **Google DeepMind**,　　Feb 2025-Now
Veo post-training: One model with the joint capabilities of Camera, Reference, Motion, I2V, T2V, etc..
Vision tokenizer: Representation learning to find a joint space encoding RGB and controls.

Research Intern, **Adobe Research**,　　May 2024-Aug 2024
Supervised by Jason Kuen and Zhe Lin
Research Topic: Autoregressive Visual Generation

Research Intern, **Microsoft**, Azure Computer Vision Team　　May 2023-Aug 2023
Supervised by Zicheng Liu, Yinpeng Chen, Chung-Ching Lin and Lijuan Wang
Research Topic: Bridging Object Generation and Segmentation

Research Intern, **Tiktok**, Multimedia Lab　　May 2022-Aug 2022
Supervised by Shijie Zhao and Junlin Li
Research Topic: Audiovisual Salient Object Detection

Research Intern, **Microsoft Research Asia**　　Oct 2020-Aug 2021
Supervised by Jinglu Wang
Research Topic: Multimodal Video Instance/Object Segmentation

PUBLICATIONS

[**ICML 2025**] Hao Chen, Yujin Han, Fangyi Chen, **Xiang Li**, Yidong Wang, Jindong Wang, Ze Wang, Zicheng Liu, Difan Zou, Bhiksha Raj *"Masked Autoencoders Are Effective Tokenizers for Diffusion Models"*

[**CVPR 2025**] Hao Chen, Ze Wang, **Xiang Li**, Ximeng Sun, Fangyi Chen, Jiang Liu, Jindong Wang, Bhiksha Raj, Zicheng Liu, Emad Barsoum *"SoftVQ-VAE: Efficient 1-Dimensional Continuous Tokenizer"*

[**ICLR 2025**] **Xiang Li**, Hao Chen, Kai Qiu, Jason Kuen, Jiuxiang Gu, Bhiksha Raj, Zhe Lin *"ImageFolder: Autoregressive Image Generation with Folded Tokens"*

[**ICLR 2025**] Xiaohao Xu, Tianyi Zhang, Shibo Zhao, **Xiang Li**, Sibo Wang, Yongqi Chen, Ye Li, Bhiksha Raj, Matthew Johnson-Roberson, Sebastian Scherer, Xiaonan Huang *"Scalable Benchmarking and Robust Learning for Noise-Free Ego-Motion and 3D Reconstruction from Noisy Video"*

[**Arxiv 2024**] Hao Chen, Abdul Waheed, **Xiang Li**, Yidong Wang, Jindong Wang, Bhiksha Raj, Marah I Abdin *"On the Diversity of Synthetic Data and its Impact on Training Large Language Models"*

[**NeurIPS 2024**] Hao Chen, Yujin Han, Diganta Misra, **Xiang Li**, Kai Hu, Difan Zou, Masashi Sugiyama, Jindong Wang, Bhiksha Raj *"Slight Corruption in Pre-training Data Makes Better Diffusion Models"*

[**NeurIPS 2024**] Kai Hu, Weichen Yu, Tianjun Yao, **Xiang Li**, Wenhe Liu, Lijun Yu, Yining Li, Kai Chen, Zhiqiang Shen, Matt Fredrikson *"Efficient LLM Jailbreak via Adaptive Dense-to-sparse Constrained Optimization"*

[**NeurIPS 2024**] Hao Chen, Ankit Shah, Jindong Wang, Ran Tao, Yidong Wang, **Xiang Li**, Xing Xie, Masashi Sugiyama, Rita Singh, Bhiksha Raj *"Imprecise Label Learning: A Unified Framework for Learning with Various Imprecise Label Configurations"*

[**ECCV 2024**] **Xiang Li**, Kai Qiu, Jinglu Wang, Xiaohao Xu, Rita Singh, Kashu Yamazak, Hao Chen, Xiaonan Huang, Bhiksha Raj *"R^2-Bench: Benchmarking the Robustness of Referring Perception Models under Perturbations"*

[**CVPR 2024**] **Xiang Li**, Jinglu Wang, Xiaohao Xu, Rita Singh, Yan Lu, Bhiksha Raj *"Towards Robust Audiovisual Segmentation in Complex Environments with Quantization-based Semantic Decomposition"*

[**MICCAI 2024**] Yizhou Zhao, Hengwei Bian, Michael Mu, Mostofa R Uddin, Zhenyang Li, **Xiang Li**, Tianyang Wang, Min Xu *"Training-free CryoET Tomogram Segmentation"*

[**InterSpeech 2024**] Muqiao Yang, **Xiang Li**, Umberto Cappellazzo, Shinji Watanabe, Bhiksha Raj *"Evaluating and Improving Continual Learning in Spoken Language Understanding"*

[**ICML 2024**] Hao Chen, Jindong Wang, Lei Feng, **Xiang Li**, Yidong Wang, Xing Xie, Masashi Sugiyama, Rita Singh, Bhiksha Raj *"A General Framework for Learning from Weak Supervision"*

[**ICML 2024**] **Xiang Li**, Yinpeng Chen, Chung-Ching Lin, Rita Singh, Bhiksha Raj, Zicheng Liu *"Completing Visual Objects via Bridging Generation and Segmentation"*

[**NAACL 2024**] Zhaorun Chen, Zhihong Zhu, Ruiqi Zhang, **Xiang Li**, Bhiksha Raj, Huaxiu Yao *"AutoPRM: Self-supervised Fine-grained Feedback for Multi-Step Reasoning via Controllable Question Decomposition"*

[**ICASSP 2024**] Muqiao Yang, Umberto Cappellazzo, **Xiang Li**, Shinji Watanabe, Bhiksha Raj *"Improving Continual Learning of Acoustic Scene Classification via Mutual Information Optimization"*

[**ICPR 2024**] Jimin Sohn, Haeji Jung, Zhiwen Yan, Vibha Masti, **Xiang Li**, Bhiksha Raj *"Fashion Image Retrieval with Occlusion"*

[**Computer Speech & Language 2024**] Fan Yang, Muqiao Yang, **Xiang Li**, Yuxuan Wu, Zhiyuan Zhao, Bhiksha Raj, Rita Singh *"A Closer Look at Reinforcement Learning-based Automatic Speech Recognition"*

[**EMNLP 2023**] **Xiang Li**, Jinglu Wang, Xiaohao Xu, Muqiao Yang, Rita Singh, Bhiksha Raj *"Towards Noise-Tolerant Speech-Referring Video Object Segmentation: Bridging Speech and Text"*

[**NeurIPS 2023**] **Xiang Li**, Chung-Ching Lin, Yinpeng Chen, Jinglu Wang, Zicheng Liu, Bhiksha Raj *"PaintSeg: Training-free Segmentation via Painting"*

[**ICCV 2023**] **Xiang Li**, Jinglu Wang, Xiaohao Xu, Xiao Li, Bhiksha Raj, Yan Lu *"Robust Referring Video Object Segmentation via Relational Multimodal Cycle Consistency"*

[**ACM MM 2023**] **Xiang Li**, Yandong Wen, Muqiao Yang, Jinglu Wang, Rita Singh, Bhiksha Raj *"Rethinking Voice-Face Correlation: A Geometry View"*

[**InterSpeech 2023**] Liao Qu*, Xianwei Zou*, **Xiang Li***, Yandong Wen, Rita Singh, Bhiksha Raj *"The Hidden Dance of Phonemes and Visage: Unveiling the Enigmatic Link between Phonemes and Facial Features"*

[**AAAI 2023**] **Xiang Li**, Haoyuan Wang, Shijie Zhao, Junlin Li, Li Zhang, Bhiksha Raj *"Panoramic Video Salient Object Detection with Ambisonic Audio Guidance"*

[**AAAI 2022**] **Xiang Li**, Jinglu Wang, Xiao Li, Yan Lu *"Hybrid Instance-aware Temporal Fusion for Online Video Instance Segmentation"*

[**TMM 2022**] **Xiang Li**, Jinglu Wang, Xiao Li, Yan Lu *"Video Instance Segmentation by Instance Flow Assembly"*

[**Ubicomp 2022-HASCA**] **Xiang Li***, Jinqi Luo*, Rabih Younes *"ActivityGAN: Generative Adversarial Networks for Data Augmentation in Sensor-Based Human Activity Recognition"* (**Best Paper Award**)

[**TWC 2022**]Yong Huang, **Xiang Li**, Wei Wang, Tao Jiang, Qian Zhang *"Forgery Attack Detection in Surveillance Video Streams Using Wi-Fi Channel State Information"*

[**WCNC 2022**] Chenhui Zhao, **Xiang Li**, Rabih Younes *"Self-supervised Multi-Modal Video Forgery Attack Detection"*

[**INFOCOM 2021**]Yong Huang, **Xiang Li**, Wei Wang, Tao Jiang, Qian Zhang *"Towards Cross-Modal Forgery Detection and Localization on Live Surveillance Videos"*

[**IJCAI 2020-DL-HAR**] Jinqi Luo, **Xiang Li**, Rabih Younes *"Towards Synthetic Data Generation and Interpretation on Sensor-based Fine-grained Hand Activity Recognition"*

TEACHING   **Graduate Teaching Assistant at CMU**

- 11785, Introduction to Deep Learning, Fall 2021

- 18661, Introduction to Machine Learning, Fall 2023

- 18789, Deep Generative Modeling, Spring 2024

ACADEMIC   **Reviewer**
SERVICE

- Conference: CVPR, ECCV, ICCV, ACL, EMNLP, NAACL, NeurIPS, ICML, ICLR

- Journal: TIP, TMM, TWC