

Yi Li

PhD @ UCSD · Machine Learning Researcher @ Qualcomm

✉ jerry19992009@gmail.com · 🏠 jerryli.github.io · 🗣️ JerryYLi · 🌐 yi-li-jerry

Research Interests

- Multimodal & Agentic AI** Vision-language models • Video understanding • Multimodal agents
- Efficiency & Reliability** Efficient inference • Model reliability • Privacy & fairness
- Robotics & Physical AI** Spatial reasoning • Memory • Embodied intelligence

Work Experience

Qualcomm

San Diego, CA

Machine Learning Researcher

2024 - Present

- Efficient vision & multimodal foundation models on the edge
- Multimodal models and agents for video understanding, spatial reasoning, robotics, and physical AI

Intel Labs

San Diego, CA

Research Intern

Summer 2022

- **Collaborators:** Subarna Tripathi, Kyle Min
- Learning long-term spatiotemporal representations from video-language pretraining

MIT-IBM Watson AI Lab

Cambridge, MA

Research Intern

Summer 2021

- **Collaborators:** Rameswar Panda, Yoon Kim, Chun-Fu Chen, Rogerio Feris, David Cox
- Improving text-only machine translation by visual hallucination learned on image-text corpora

Statistical Visual Computing Lab, UC San Diego

San Diego, CA

Graduate Student Researcher

2018 - 2024

- Quantitative measurement and systematic mitigation of data bias
- Principled approach for collecting large-scale, bias-controlled datasets
- Model debiasing for improved performance, transferability and robustness in video action recognition

Education

University of California San Diego

San Diego, CA

PhD in Electrical & Computer Engineering

2017 - 2024

- **Specialization:** Signal & image processing
- Joined Statistical Visual Computing Lab (SVCL) in 2017, conducting research in computer vision
- **Thesis:** Holistic Bias Mitigation in Computer Vision and Beyond
- **Advisor:** Prof. Nuno Vasconcelos, **GPA:** 3.97/4.

- **Specialization:** Digital signal processing (DSP) and multimedia
- Entered ELITE Stream of the Faculty of Engineering in 2015
- **Thesis:** Binary Image Denoising with BinURE-LET
- **Advisors:** Prof. Thierry Blu & Prof. Tan Lee, **Major GPA:** 4.00/4.

Publications

Conditional Multi-Event Temporal Grounding in Long-Form Video

NeurIPS 2026 E&D

Yuanhao Zou, Arthad Kulkarni, Lucas Tonanez, Lincoln Spencer, Guangyu Sun, Tianxingjian

Ding, Andong Deng, Yi Li, Shuangjun Liu, Yuan Li, Dashan Gao, Ning Bi, Taotao Jing, Shuai

Atlanta, GA

Zhang, Chen Chen

- Introduced CoMET-Bench for conditional multi-event grounding, counting, and negative-query recognition in long-form video
- Proposed CoMET-Agent, a training-free framework for structured search and aggregation

ForeSea: AI Forensic Search with Multi-modal Queries for Video Surveillance

ECCV 2026

Hyojin Park*, Yi Li*, Janghoon Cho, Sungha Choi, Jungsoo Lee, Taotao Jing, Shuai Zhang,

Munawar Hayat, Dashan Gao, Ning Bi, Fatih Porikli

Malmö, Sweden

- Introduced ForeSeaQA, a benchmark for multimodal video question answering and temporal grounding
- Developed a retrieval-augmented video search pipeline combining tracking, multimodal indexing, and video-language reasoning

Leveraging Data to Say No: Memory Augmented Plug-and-Play Selective Prediction

ICLR 2026

Aditya Sarkar, Yi Li, Jiacheng Cheng, Shlok Mishra, Nuno Vasconcelos

Rio de Janeiro, Brazil

- Proposed memory-augmented plug-and-play selective prediction for vision-language foundation models
- Used retrieved image-text pairs and contrastive normalization to improve embedding stability and score calibration

EgoPrivacy: What Your First-Person Camera Says About You?

ICML 2025

Yijiang Li, Genpei Zhang, Jiacheng Cheng, Yi Li, Xiaojun Shan, Dashan Gao, Jiancheng Lyu,

Yuan Li, Ning Bi, Nuno Vasconcelos

Vancouver, Canada

- Proposed EgoPrivacy, a large-scale benchmark of wearer privacy in first-person (egocentric) videos
- Studied egocentric privacy risks at demographic, individual, and situational levels with various attack models
- Introduced new privacy attack through ego-to-exo view retrieval augmentation

SViTT: Temporal Learning of Sparse Video-Text Transformers

CVPR 2023

Yi Li, Kyle Min, Subarna Tripathi, Nuno Vasconcelos

Vancouver, Canada

- Proposed SViTT, a sparse video-text architecture with hybrid edge and node sparsity
- SViTT is trained with a temporal expansion curriculum with increasing clip length and model sparsity
- SViTT outperforms dense transformers on video retrieval and question answering, with reduced compute cost

VALHALLA: Visual Hallucination for Machine Translation

CVPR 2022

Yi Li, Rameswar Panda, Yoon Kim, Chun-Fu Chen, Rogerio Feris, David Cox, Nuno Vasconcelos

New Orleans, LA

- Proposed VALHALLA framework for improving natural language modeling from multimodal data
- Trained on image-text corpora, VALHALLA learns to hallucinate visual features to assist text-only inference
- Improved machine translation results on text-only & multimodal datasets over state-of-the-art

Improving Video Model Transfer with Dynamic Representation Learning

CVPR 2022

Yi Li, Nuno Vasconcelos

New Orleans, LA

- Proposed DRL, a framework for learning video representations that capture temporal dynamics
- By maximizing their dynamic score, DRL prevents networks from overfitting to spatial bias in video data
- Improved transferability and generalization of video models over conventional training

BEV-Net: Assessing Social Distancing Compliance by Joint People Localization and Geometric Reasoning

ICCV 2021

Zhirui Dai, Yuepeng Jiang, Yi Li, Bo Liu, Antoni B. Chan, Nuno Vasconcelos

Virtual

- Proposed BEV-Net, a framework for localizing people in world frame from monocular cameras
- Developed a differentiable homography module to map image into BEV coordinates for end-to-end training
- Released a new benchmark on social distancing compliance assessment

Learning Representations from Audio-Visual Spatial Alignment

NeurIPS 2020

Pedro Morgado*, Yi Li*, Nuno Vasconcelos

Virtual

- Proposed audio-visual spatial alignment (AVSA) task for self-supervised learning from 360° videos
- AVSA uses contrastive learning to match audio and visual signals from the same viewpoints
- Developed a transformer-based architecture for AVSA which models context across viewpoints
- Showed that training with AVSA improves representation for various downstream tasks

Background Data Resampling for Outlier-Aware Classification

CVPR 2020

Yi Li, Nuno Vasconcelos

Virtual

- Proposed a resampling approach for out-of-distribution (OOD) detection with background data
- Formulated background dataset selection as adversarial reweighting problem
- Achieved superior OOD detection with resampled background data vs. uniform sampling

REPAIR: Removing Representation Bias by Dataset Resampling

CVPR 2019

Yi Li, Nuno Vasconcelos

Long Beach, CA

- Developed a differentiable data resampling approach (REPAIR) to reduce representation bias in datasets
- Applied the proposed method on action recognition datasets to remove static bias from videos
- Evaluated sensitivity of video recognition algorithms to the presence of bias in data
- Showed removing representation bias improves model generalization

RESOUND: Towards Action Recognition without Representation Bias

ECCV 2018

Yingwei Li, Yi Li, Nuno Vasconcelos

Munich, Germany

- First study of representation bias in video action recognition datasets
- Quantitatively measured the object/scene/human bias of existing action recognition datasets
- Collected Diving48, a dataset of competitive diving for bias-free action recognition

* Equal contribution.

Academic Service

Peer reviewer TPAMI, CVPR (**top reviewer** '23), ICCV, ECCV, NeurIPS (**top reviewer** '23), ICLR
Student volunteer CVPR '19/20 AC meeting

Honors & Awards

2017 **Graduate Fellowship**, Dept. of Electrical & Computer Engineering, UCSD
Winbridge Scholarship, CW Chu College, CUHK
2016 **HKSAR Government Scholarship**, The Government of Hong Kong
EE Alumni Association Scholarship, Dept. of Electronic Engineering, CUHK

Teaching Experience

Teaching Assistant ECE 153 Probability & Random Processes for Engineers
ECE 271A Statistical Learning I
ECE 271B Statistical Learning II
ECE 271C Deep Learning and Applications

Skills

Programming Languages Python, MATLAB, C/C++, Java
Deep Learning Libraries PyTorch, JAX/Flax, TensorFlow, Caffe
Miscellaneous $\LaTeX$, HTML/CSS, Bash
Languages English, Chinese