

Qian Jiang

✉ qianj3@illinois.edu 🌐 qianjiangcn.github.io 📡 qianj3 💬 qianjiangcn

Education

Ph.D. in Electrical and Computer Engineering <i>University of Illinois at Urbana-Champaign (UIUC)</i> Advisor: Professor Minh N. Do	2019- 2024 <i>Illinois, USA</i>
B.Sc. in Electrical and Computer Engineering <i>University of Electronic Science and Technology of China (UESTC)</i>	2015 - 2019 <i>Chengdu, China</i>

Research Interests

- Computer Vision, Large Language Models(LLMs), Multi-modal Learning, Generative Modeling, and Efficient Machine Learning.

Work Experience

AWS AI Labs <i>Applied Scientist, LLM Applications</i>	2024 - Now <i>Santa Clara, USA</i>
● Working on LLM-assisted Code Development. ● Working on architect and implement advanced features for code documentation and migration platform, enabling automated code analysis and transformation. ● Expertise with state-of-the-art large language models.	
Microsoft <i>Research Intern, Multimodal Generative Modeling</i>	2023 Fall <i>Redmond, USA</i>
● Worked diffusion-based modeling with extensions into text-controlled video generation. ● Developed efficient text-to-video fine-tuning approaches.	
Amazon <i>Applied Scientist Intern, Large-scale Models for Relevance Matching</i>	2023 Summer <i>Seattle, USA</i>
● Developed methods for Amazon product relevance matching and improved 4% on ranking metric. ● Developed stochastic approach for contrastive learning on public benchmarks (CC3M, ImageNet) with improved performance over SOTA. ● Two papers submitted.	
Amazon <i>Applied Scientist Intern, Large-scale Vision-Language Representation Learning</i>	2022 Summer Fall <i>Seattle, USA</i>
● Conducted empirical and theoretical analysis on modality alignment. ● Propose three instrumental regularizations to improve latent modality structures. ● Conduct extensive and comprehensive experiments on various vision-language models for different downstream applications (e.g., cross-modality retrieval, VQA, VR and etc). ● One paper accepted to CVPR'23.	
IBM Research <i>Research Intern, Optimization of parallelization for IBM clouds</i>	2020 Summer <i>Yorktown Heights, USA</i>
● Developed tools to efficiently benchmark and visualize parallel communication performance. ● Optimized parameters for message passing interface on IBM clouds.	

Research Experience

University of Illinois at Urbana-Champaign (UIUC)

2019 - 2024

Retrieval Augmented Generation

Illinois, USA

- Developed GPTFedRec, a federated recommendation framework leveraging ChatGPT and a novel hybrid Retrieval Augmented Generation (RAG) mechanism.
- Conducted experiments on multiple recommendation tasks with improved performance over SOTA.

Neural Architecture Search

- Developed differentiable models predicting end-to-end hardware performance of neural network architectures.
- Conducted experiments on CIFAR and ImageNet datasets on multiple hardware platforms (Edge GPUs, Edge TPUs, Mobile CPUs, and customized accelerators) with improved performance.

Multi-source Transfer Learning

- Formulated multi-source transfer learning as a bi-level optimization problem.
- Conducted experiments on multiple tasks including classification and scene understanding.

Review Services

- Conferences: NeurIPS, ICML, ICLR, ICASSP.

- Journals: Pattern Recognition.

Publications

GenAI, Multi-Modal, Foundation Models

1. **Qian Jiang**, Changyou Chen, Han Zhao, Liqun Chen, Qing Ping, Son Dinh Tran, Yi Xu, Belinda Zeng, Trishul Chilimbi. "Understanding and Constructing Latent Modality Structures in Multi-modal Representation Learning." CVPR'23. [\[Paper\]](#)
2. Huimin Zeng, Zhenrui Yue, **Qian Jiang**, Dong Wang. "Federated Recommendation via Hybrid Retrieval Augmented Generation." IEEE BigData'24. [\[Paper\]](#)
3. **Qian Jiang**, Jingjing Meng, Alireza Bagheri Garakani, Yang Jiao, Yetian Chen, Yikai Ni, Yan Gao, Yi Sun, Changyou Chen. "SLA: Stochastic Label Augmentation for Robust Vision-Language Contrastive Learning." Under Review.

Efficient ML

1. **Qian Jiang***, Xiaofan Zhang*, Deming Chen, Minh N. Do, and Raymond A. Yeh. "EH-DNAS: End-to-end hardware-aware differentiable neural architecture search." ICML'23 Workshop on Differentiable Almost Everything. [\[Paper\]](#) [\[Code\]](#)
2. **Qian Jiang**, Raymond A. Yeh, and Minh N. Do. "Multi-source transfer learning by learning to weight source tasks." Under Review.

Teaching Experience

University of Illinois at Urbana-Champaign

2021-2022

Teaching Assistant, Electrical and Computer Engineering Department

Illinois, USA

- ECE310: Digital Signal Processing
- ECE311: Digital Signal Processing Lab

Other Experience

University of California, Los Angeles (UCLA)

2018 Summer

Full scholarship, Cross-disciplinary Scholar in Science and Technology (CSST)

Los Angeles, UCLA

Scholarships and Awards

- Nadine Barrie Smith Memorial Fellowship, 2022.
- C3SR (Illinois- IBM Center of Cognitive Computing Systems Research) Fellowship, 2019-2021.
- National Scholarship, China, 2018.
- Tanglixin Scholarship, China, 2017.

Programming Skills

- Pytorch, Python, MATLAB, Bash, Vim, Git.

Relevant Coursework

- **Optimization and Statistics:** Random Processes; Introduction to Optimization; Statistical Learning Theory; Computational Inference; Information Theory; Games, Markets, and Mathematical Programming.
- **Machine Learning:** Pattern Recognition; Computer Vision; Deep Generative and Dynamical Models; Mathematical Models of Language.