Yu (Roy) Liu

1 Ely Park Blvd Apt 24-4, Binghamton, NY, 13905
yliu456@binghamton.edu, 704-8587806
https://github.com/royliuyu

SUMMARY

Extensive expertise in machine learning, demonstrated by a proven track record of developing and optimizing state-of-the-art algorithms, coupled with practical experience in data engineering. Proficient in computer vision and large language models (LLMs), with a strong foundation in computer science principles such as computer architecture, databases, and system scheduling, enabling efficient data optimization, algorithm refinement, and resource management.

Distinguished by research accomplishments during a Ph.D., including the publication of five first-author SCI papers. Active engagement with the AI community has led to top rankings in Kaggle competitions, achieving placements in the top 2.6% and 4.9% (silver prize) in two recent challenges.

Skilled in cross-functional collaboration to drive innovation and address complex real-world challenges. A commitment to staying at the cutting edge of AI advancements ensures the delivery of impactful and forward-thinking solutions in dynamic environments.

EDUCATION

State University of New York at Binghamton — **Jan. 2021 – Dec. 2024**
Ph.D, Computer Science — GPA: 3.8/4.0

University of North Carolina at Charlotte — **Aug. 2019 – Dec. 2020**
Ph.D, Electrical Computer Engineering — GPA: 4.0/4.0

Shanghai Jiao Tong University — **Sep. 1998 – Jan. 2001**
Master, Business Administration — GPA: 3.4/4.0

Shanghai University of Engineering Science — **Sep. 1990 – Jul. 1994**
Bachelor, Computer and Electronics — GPA: 3.2/4.0

PROFESSIONAL EXPERIENCES

Research Foundation for State University of New York - Research Assistant — **Feb. 2021 – Present**
Research Assistant

- Investigate recent advancements and innovations, defining research objectives, topics, and timelines
- Generate novel ideas and solutions in key research areas, such as 3D object detection/tracking, efficient edge-service system design, and enhancing LLM inference throughput
- Perform data engineering using large datasets such as Azure Trace and NuScenes, along with annotation and labeling of proprietary datasets like self-recorded videos
- Design, build, train, and optimize advanced AI networks; conduct experiments to validate proposed methods and publish findings
- Foster strong connections with academic and professional institutions to stay engaged in the field. For example, I attended and presented at the 100th IEEE Vehicular Technology Conference in October 2024, where I also chaired a conference session
- Actively contribute to the AI community by regularly participating in Kaggle competitions, leveraging state-of-the-art networks and algorithms, including machine learning, computer vision models, and LLMs, to tackle real-world challenges across a variety of datasets and domains

Shanghai General Grid Technology Co., Ltd. — **Sep. 2012 – Jul. 2019**
Founder

- Spearheaded the R&D efforts to develop the "Intelligent Parcel Locker". Direct the embedded systems team to create the Locker Control Unit module using MCU-based technology and develop a proprietary dispatch management system built on a client-server (C/S) architecture. The cloud system integrates WINFORM, WCF, and SQL, with web and mobile applications on the client side. Secure nine software copyrights for these innovations
- Strategically collaborate with cross-functional teams to expand market penetration of the product across diverse industries, including e-commerce and utilities, both domestically and internationally

General Electric Company, GE Power — **Aug. 2005 – Nov. 2011**
Asia Pacific Operation Manager

- Lead the development and deployment of "Operations and Product Management systems", globally adopted by GE Energy, leveraging a B/S framework within ASP.NET+SQL architecture
- Optimize supply chain operations across Asian factories by centralizing SKU, cost, and pricing data into a unified database, enhancing operational efficiency and cost-effectiveness

Fuji Xerox (China) Co., Ltd. — **Aug. 2000 – Aug. 2005**
Operation Manager

- Propose one-step office solutions to clients by integrating with products of multifunction printers and software, delivering enhanced productivity and efficiency.
- Direct sales operations, manage RFQs, optimize and implement effective sales policies to drive business growth

ABB (China) Co., Ltd.   **Aug. 1997 – Aug. 2000**
Technical Engineer
- Support the design of electrical systems with a wide range of electrical products, contributing to successful project implementations and client satisfaction

PUBLICATIONS
- Yu Liu and KD Kang. "AROD: Adaptive Real-Time Object Detection Based on Pixel Motion Speed", IEEE 100th Vehicular Technology Conference (VTC2024-Fall), 2024
- Yu Liu, Anurag Andhare, and Kyoung-Don Kang. "Corun: Concurrent Inference and Continuous Training at the Edge for Cost-Efficient AI-Based Mobile Image Sensing", Sensors 24(16): 5262, 2024
- Yu Liu and KD Kang. "Filtering Empty Video Frames for Efficient Real-Time Object Detection", Sensors 24(10): 3025, 2024
- Yu Liu and KD Kang. "Preprocessing via Deep Learning for Enhancing Real-time Performance of Object Detection", IEEE 97th Vehicular Technology Conference (VTC2023-Spring), pages 1–5. IEEE, 2023
- Yu Liu, KD Kang, and Mi Jin Doe. "HADD: High-accuracy Detection of Depressed Mood". Technologies, 10(6):123, 2022

SKILLS
- Machine Learning & AI: Proficient in traditional ML algorithms as well as popular AI architectures, including LLMs, transformers, CNNs, GANs, and diffusion models. Skilled in developing and training advanced AI algorithms. Experienced in utilizing various AI techniques, such as attention mechanisms, transfer learning, and parallelism, to achieve state-of-the-art performance in both accuracy and efficiency
- Computer system expertise: Extensive experience with client-server and edge-server architectures, database systems, as well as system scheduling, multiprocessing, and multithreading mechanisms
- Data Science: Extensive knowledge across multiple areas of data science, including data cleaning, data engineering, ensemble modeling, hyper-parameter tuning, feature extraction, and model optimization
- Programming: Fluent in Python, with additional capabilities in C and C++
- Cross-functional collaboration: Substantial working experience across multiple functions in Business Operations, including strategic planning, process optimization, and workflow management

PROJECTS & OUTSIDE EXPERIENCE
Real-time 3D Object Detection for Autonomous Driving with Lidar and Camera Data   **Dec. 2023 – Present**
- Conduct a comprehensive literature survey of related work and investigate existing state-of-the-art algorithms
- Design the novel network that integrates transformer models, sparse voxels, and temporal attention techniques to improve 3D object detection efficiency without increasing latency

Efficient Super Resolution project   **Jun. 2024 – Nov.2024**
- Design a network incorporating spatial attention, bidirectional channel attention, and depth-wise convolution techniques to enhance efficiency without compromising accuracy
- The new model achieves only 20% of the baseline's parameter size and GFLOPs

LLM Prompt Recovery competition   **Feb. 2024 – Apr. 2024**
- Recover the LLM prompt used for rewriting texts, assessed using a dataset of 1,300+ original texts paired with versions
- Develop code to fine-tune the pre-trained Gemma-7b, Phi-2, and Mistral-7b models to generate the prompt in reverse

Preprocess video via DL for Enhancing Real-time Performance of Object Detection   **Jun. 2023 – May 2024**
- Design a CNN-LSTM hybrid network to improve efficiency in video object detection without sacrificing accuracy
- Collect and process dataset, and conduct experiments to validate performance and effectiveness
- Published papers detailing the network and findings

Enhance the CNN Inference Throughput via Concurrent Model Training and Inferences   **Aug. 2023 – Feb.2024**
- Independently develop a client-server AI serving platform capable of executing multiple inference and training processes. The system supports a wide range of CNN and transformer models
- Efficiently manage CPU-GPU collaboration and implement GPU resource partitioning techniques (e.g., Nvidia MPS) to enhance throughput
- Published a paper

HOBBIES & INTERESTS
- Skiing, billiards, swimming, and reading