Shang-Jui (Ray) Kuo

AI Researcher, Inventec Corp., Taiwan raykuo.sj@gmail.com Personal Website

Research Interests

With research experience in few-shot learning, vision-language models, and algorithm-hardware co-design, I've focused on enhancing data efficiency and inference efficiency. Now, I'm eager to leverage this expertise, integrating knowledge distillation techniques, to downsize vision and language models for specialized domains, contributing to the advancement of efficient and deployable AI solutions.

Education

B.S. Electrical Engineering, National Taiwan University (NTU) **Taipei, Taiwan**

Overall GPA: 3.62/4.3, Last-60-Credits Average: 4.02/4.3 *Sep. 2019 - June 2023*

- **Artificial intelligence:** Deep Learning for Computer Vision*, Machine Learning*, Probability and Statistics, Linear Algebra, Differential Equation, Algorithms, Data Structure, Signals and Systems
- **Domain knowledge:** Introduction to Cryptography, Introduction to Information Security, Electrical Engineering Lab-Networking and Multimedia, Electrical Engineering Lab-Embedded System, Introduction to Computer Networks
- **Advanced Hardware design:** Digital Signal Processing in VLSI Design*, Digital System Design, Integrated Circuit Design, Integrated Circuits Design Laboratory, Computer Architecture (*denotes graduate level)

Publication

Improving Limited Supervised Foot Ulcer Segmentation Using Cross-Domain Augmentation - LINK

Shang-Jui Kuo*, Po-Han Huang*, Chia-Ching Lin, Jeng-Lin Li, Ming-Ching Chang

Accepted by 2024 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2024).

Patent

Segmentation Model Training Method, Device, and Non-transitory Computer Readable Storage Medium (Pending)

Shang-Jui Kuo, Po-Han Huang, Chia-Ching Lin, Jeng-Lin Li, Ming-Ching Chang, Wei-Chao Chen

Awards

Won First Place in 2023 Inventec Hackathon Competition (Team Leader) **Taipei, Taiwan**

Artificial Intelligent 5G Beamforming Antenna System *Oct. 2023*

- Led a five-member team to achieve both first place and the Best Presentation Award at the competition in a highly competitive event hosted by Inventec Corporation, featuring over 300 participants and 74 teams.
- Integrated beamforming technology and AI to create a versatile prototype, featuring an AI model for detecting individuals, identifying moving objects, and determining device orientation.

Research Experience

AI Researcher at Inventec Corporation, AI Center, Research Team **Taipei, Taiwan**

Supervisor: Dr. Wei-Chao Chen, Prof. Ming-Ching Chang *July 2023 – Present*

Improving Limited Supervised Foot Ulcer Segmentation Using Cross-Domain Augmentation

- Developed a novel methodology to enhance limited wound segmentation using cross-domain augmentation strategies, and the results are currently under review for publication at ICASSP-2024.
- Realized a 34% improvement in the Dice score for a real-world segmentation project by implementing the developed method in collaboration with Inventec Corporation and Taipei Veterans General Hospital.
- Submitted a patent application based on our research findings, contributing to the company's competitive edge.

Student Researcher at National Taiwan University, Vision & Learning Lab **Taipei, Taiwan**

Supervisor: Prof. Yu-Chiang Frank Wang *Aug. 2022 – June 2023*

Text-to-Image (T2I) Synthesis & Image Manipulation

- Introduced an innovative approach in text-to-image synthesis, incorporating images as guidance to address limitations in precise image control solely through textual inputs.
- Proposed a novel concept for leveraging synthetic datasets to improve the performance of vision-language models in countering challenges associated with counting tasks.

Student Researcher at National Taiwan University, Media IC & System Lab **Taipei, Taiwan**
Supervisor: Prof. Shao-Yi Chien *Aug. 2021 – June 2022*

Researched and enhanced accelerator efficiency

- Implementing a sparse-aware DNN hardware accelerator on Altera DE2-115 FPGA, which successfully executes real-time 720p Visual Object Tracking (VOT) algorithms.
- Utilizing algorithm-hardware co-design techniques, pruned the VOT model to reduce 70% of parameters and designed an advanced sparse-aware DNN hardware accelerator to maximize throughput and lower latency.
- Developed a tool for modeling and evaluating the performance and energy efficiency of different dataflows, optimally mapping each layer in the VOT model to our hardware accelerator.

Professional Services

Journal Reviewer
Computers & Graphics (2023)
ACM Transactions on Multimedia Computing, Communications, and Applications

Work and Teaching Experience

Guest Lecturer for Embedded Deep Neural Network Processing Course **Taipei, Taiwan**
National Taiwan University of Science and Technology (Taiwan Tech) *Dec. 2023*

- Delivered a session on my experience with the AIoC team at Inventec Corp., including model training for resource-limited hardware, model quantization, and hardware verification.
- Delivered a session on the topic of AI accelerators, covering deep neural network hardware accelerator architecture design, dataflow simulation tools, and algorithm-hardware co-design.

AI Accelerator Engineer at Inventec Corporation, Digital Center, AI on Chip Team **Taipei, Taiwan**
Supervisor: Dr. Wei-Chao Chen *July 2023 - Present*

- Improved inference efficiency via flexible accelerator-CPU collaboration, streamlining computations, and simplifying circuits, especially for challenging operations like Softmax, non-maximum suppression, and keypoint alignment.
- Developed a co-simulation NPU simulator and testing environment to accelerate AI model deployment on hardware.
- Developed a visualization tool for hardware verification results to assist the verification process.

Intern at Inventec Corporation, Digital Center, AI on Chip Team **Taipei, Taiwan**
Supervisor: Dr. Wei-Chao Chen *March 2023 - June 2023*

- Designed a neural network model for a top IC design house, embedded in the panel IC for stylus trajectory prediction.
- Achieved a noteworthy 50% error reduction with my model, heightening the product's competitive advantage.

Selected Projects

Contributor of Apache Open-Source Project: Apache Submarine - LINK **Sept. 2021 – June 2022**

- Participated actively in the Apache Submarine project, a cloud-native machine learning platform for ML engineers.
- Demonstrated versatile contributions spanning front-end UI, and back-end server, ensuring compatibility across various CPU architectures, and configuring development environments for a seamless user experience.

Technical Skills and Language

Languages: TOEFL iBT: 95, GRE: Verbal-157, Quant-168, AW-3.0
Programming Languages: Python, C++, TypeScript, JavaScript, Verilog, SystemVerilog
Technologies/Frameworks: Pytorch, Quantization Aware Training (AIMET Model Quantization, OpenVINO)
Developer Tools: Git, Docker, Kubernetes, EDA tool (Vivado, NC-Verilog, Design Compiler)
Evaluation Board: Altera DE2-115 FPGA, Raspberry Pi, STM32

Extracurricular and Leadership

NTU Men's Basketball Varsity Team **Taiwan**
Starter *July 2020 - March 2021*

- As a starting player, I led the team from Division 2 to Division 1, making it the only Division 1 team in the history of Taiwan without any basketball-specialized students.

References

Wei-Chao Chen
Chief Digital Officer
Senior Vice President
Inventec Corporation
chen.wei-chao@inventec.com

Ming-Ching Chang
Assistant Professor
Department of Computer Science
College of Engineering and Applied Sciences
University at Albany, State University of New York
mchang2@albany.edu

Yu-Chiang Frank Wang
Professor
Department of Electrical Engineering
National Taiwan University
Director of Computing Center, Dept. EE
ycwang@ntu.edu.tw

Shao-Yi Chien
Professor
Department of Electrical Engineering
National Taiwan University
sychien@ntu.edu.tw