Shang-Jui (Ray) Kuo

+1-934-246-1740 | raykuo.sj@gmail.com | raykuo18.github.io | github.com/raykuo18

Education

Stony Brook University — Stony Brook, NY
Ph.D. in Computer Science – Advisor: Prof. Paola Cascante-Bonilla, SPELL Lab — *Aug. 2024 – Present*

National Taiwan University — Taipei, Taiwan
B.S. in Electrical Engineering – GPA: 3.62/4.3 | Last 60 credits: 4.02/4.3 — *Sept. 2019 – June 2023*

Publications

Do VLMs Need Vision Transformers? Evaluating State Space Models as Vision Encoders

- **Shang-Jui (Ray) Kuo**, Paola Cascante-Bonilla. *Under Review*, 2026. arXiv | Project Page | Code | HF Paper | Checkpoints – Featured on Hugging Face Daily Papers

Improving Limited Supervised Foot Ulcer Segmentation Using Cross-Domain Augmentation

- **Shang-Jui Kuo***, Po-Han Huang*, Chia-Ching Lin, Jeng-Lin Li, Ming-Ching Chang. *ICASSP 2024*. arXiv

Research Experience

Graduate Researcher – SPELL Lab — Sept. 2025 – Present
Stony Brook University | Advisor: Prof. Paola Cascante-Bonilla — *Stony Brook, NY*

- Designed a controlled study comparing SSM (VMamba, MambaVision), ViT, and hybrid vision encoders as frozen backbones in a fixed LLaVA-style pipeline, questioning whether ViTs are the right default for VLMs
- Found SSM backbones deliver stronger localization while remaining competitive on open-ended VQA; showed ImageNet accuracy does not reliably predict downstream VLM performance
- Released model checkpoints on HuggingFace; paper featured on Hugging Face Daily Papers; poster accepted to SUNY AI Symposium 2026

Research Assistant – COMPAS Lab — Fall 2025
Stony Brook University | Advisors: Prof. Mike Ferdman, Prof. Peter Milder — *Stony Brook, NY*

- Brought up AMD AI Engine (AIE) on VCK5000 FPGA for MLISA, a hardware-agnostic ML inference framework built on LLVM IR as an alternative to the PyTorch+CUDA stack
- Navigated severely broken toolchain with no prior lab experience on this board: early-access-only documentation, non-functional hardware emulation, and silent failures
- Invented RTL-level debug methodology using configurable debug registers under 8-hour bitstream compile cycles with no simulator; successfully ran first matmul kernel on AIE in the lab

Work Experience

AI Researcher & AI Accelerator Engineer — Mar. 2023 – Feb. 2024
Inventec Corporation – AI Center & Digital Center — *Taipei, Taiwan*

- Published ICASSP 2024 paper on cross-domain augmentation for medical image segmentation; achieved 34% Dice score improvement in real hospital deployment (Taipei Veterans General Hospital)
- Designed a lightweight CNN (10–50K params) for stylus trajectory regression from capacitive sensors, achieving sub-2-pixel accuracy and ∼50% error reduction over prior rule-based approach; deployed in panel IC of a top IC design house
- Contributed to VectorMesh™ NPU IP: designed RTL image resize DSP module and NPU co-simulation testing environment; product later won EETimes 2024 Asia Golden Award

Awards & Service

1st Place + Best Presentation | Inventec Hackathon 2023 (300+ participants, 74 teams) — Oct. 2023

- Led 5-member team; built AI-powered beamforming system repurposing existing laptop antennas for human sensing and smart beam steering with no hardware modifications required

Reviewer | Computers & Graphics; ACM Trans. on Multimedia Computing (2023)