Jingtao Li

480-738-3855 • zlijingtao@gmail.com • linkedin.com/in/lijingtaoz • github.com/zlijingtao • bit.ly/jingtaoscholar

SUMMARY

Research Scientist specializing in AI privacy & security, computer vision, federated learning, and efficient algorithm design.

EDUCATION

Ph.D. Electrical Engineering August 2023
Arizona State University, Tempe, AZ 4.0 GPA

B.S. Microelectronic Science and Engineering June 2018
University of Electronic Science and Technology of China, Chengdu, China 3.9 GPA

TECHNICAL SKILLS

Programming Languages: Python, Java, C/C++, Bash
Frameworks: PyTorch, Tensorflow, TVM, Scikit-Learn
Tools and OS: Docker, AWS, Hadoop, Synopsys, Gem5, Git, SQL, MATLAB, Windows, Linux

PROFESSIONAL EXPERIENCE

Sony Corporation of America, Tempe (Remote), US: Research Scientist Aug 2023 – Now
Responsible for conducting research and landing projects in the field of private-preserving machine learning.

Arizona State University, Tempe, AZ: Graduate Research Associate Aug 2018 – May 2023
Responsible for evaluating the performance of a cutting-edge cache-reconfigurable multi-core processor.

- Developed and optimized workloads including CV workloads such as **Point Cloud**, ML workloads such as **Graph NN**, **ResNet**, and **XGBoost**, and graph algorithms such as **PageRank**, **SSSP**, and **BFS**. We achieve 69x-80x energy efficiency on assigned workloads compared to C++/CUDA implementations.
- Proposed a decision-tree-based cache reconfiguration search engine on accelerating CNN workloads, presented at ISCAS-22, and a point sampling heuristic to accelerate point cloud workloads, which was presented at SIPS-22.

Sony AI, Tokyo, JP: Research Intern May 2022 – Aug 2022
Responsible for conducting research and developing solutions in the field of private-preserving machine learning.

- Built a practical **self-supervised federated learning** system that achieves accurate and budget-friendly cross-device learning performance. Demonstrated on Raspberry Pi with 1,000 clients, which was presented at ICLR-2023 (top 5%).
- Delivered a **multi-task learning** framework for hosting multiple CV tasks given a tight memory requirement.

RESEARCH PROJECTS

Privacy & Efficiency of Federated Learning Systems Aug 2021 – Now
Paper Publication: AAAI-24, ICLR-23, CVPR-22, SIPS-21, JSPS

- Developed an "adversarial training + transfer learning" framework named ResSFL to improve the privacy of **Split Federated Learning** models. Successfully mitigated the data privacy threat of SFL, making attacks >10 times harder to succeed, with only a <1% drop in model accuracy. ResSFL is further extended to **Self-supervised Learning** as Target-aware ResSFL which showcases superior transferability.
- Designed a loss-based asynchronous **Split Federated Learning** that smartly updates model to reduce communication by up to 106.7x and computation by up to 32.1x.
- Investigate **model extraction & IP protection** of Split Federated Learning system.

Mitigating Security Issues of Edge AI System Mar 2019 – July 2021
Paper Publication: TPAMI, HOST-21, DATE-21, DAC-20, CVPR-20, SIPS-19

- Invented the first targeted **bit-flip attacks** that causes a significant accuracy drop with a few bit-flips on DNN weights. Developed RADAR, WRecon, Piece-wise Clustering defensive methods against bit-flip attacks, recovering ImageNet accuracy from <1% to above 69% with a latency overhead of <0.6% in inference shown by **Gem-5**.
- Proposed NeurObfuscator, an end-to-end DNN obfuscation framework against **side-channel attacks** targeting architecture intellectual property. Optimized obfuscation knobs using **Genetics Algorithm** and successfully defended a CTC-**LSTM**-based model architecture thief with only 2% time overhead and non-drop accuracy.

Building Energy Efficient AI System Aug 2018 – Feb 2019

Paper Publication: JETCAS, JSPS, ISCAS-22, SIPS-22

- Proposed MAX^2, an in-memory computing **AI accelerator** based on ReRAM, that maximizes data reuse and reduces on-chip bandwidth, improving computation efficiency by 2.5x and energy efficiency by 5.2x.
- Proposed decision-tree-based cache reconfiguration search engine based on a low-power **reconfigurable architecture** on accelerating CNN workloads. Proposed a point-sampling heuristic to accelerate point-cloud workloads.
- Designed an "energy + loss-aware" selective updating Learning scheduler for **energy-harvesting IoT devices**. with 8-bit float quantization to reduce communication and save energy by 43.7%-80.5% for VGG11 and models.

Design of Analog-to-Digital Converter and Semiconductor Device Apr 2016 – June 2018

Paper Publication: TCAS-I, Access, APL-a, APL-b

- Proposed ordering technique to achieve a 2~3 bits increase in the static linearity performances with elements sorting and optimal selection. Presented a statistics-optimized element organization technique in a 14-bit SAR ADC and achieved significant improvement of around 23 dB in SFDR.
- Investigated non-piezoelectric behavior at specific boron compositions and potential applications in semiconductor device development. And determined the band alignment of a β-Ga2O3/AlN heterojunction, establishing it as a type II staggered-gap heterojunction, which could enhance the design of optical and electronic devices

PATENTS

- Systems and Methods for a Full-Stack Obfuscation Framework to Mitigate Neural Network Architecture Thief (Under Provisional Application: 63/350,765)
- System and Method for Robust Neural Networking Via Noise Injection (Under Provisional Application: 17/932,104)
- Method of Arranging Capacitor Array of Successive Approximation Register Analog-to-Digital Converter (Application Granted: US10298254B1)

OTHER PROJECTS

- Job Salary Prediction (JPS-LKM) - 3rd place on the Kaggle leaderboard
- Implemented a Split Federated Learning framework on a microcontroller that has only 256KB SRAM. The demo on keyword spotting shows better accuracy than FedAvg.
- Built an expandable AR framework on Android, RANSAC pose estimation is optimized using OpenCV.
- Extended secure computing CrypTen framework with functionality support for Meanshift Clustering.
- Designed and built a Convolutional & Average Pooling engine in digital circuits, including synthesis, Layout, and post-layout evaluation with the use of Synopsys Design Compiler, Virtuoso Layout, and Primetime.

COURSES

- Secure ML computation • Embedded ML system • Statistical ML • Deep Learning
- Physics-based CV • Mobile System & Architecture • Computer Architecture • VLSI design

AWARDS

- Dean's Dissertation Award
- 57th DAC Young Fellows Poster Presentation Award
- Engineering Graduate Fellowship Award
- UESTC Outstanding Undergraduate Student Award

SERVICES

- Peer Journal Reviewer of IEEE TPDS, IEEE ESL, IEEE TGCN, IEEE TCSVT, IEEE JETCAS, IEEE TPAMI
- Conference Reviewer of ICML (2024), ICLR (2024), CVPR (2024), AAAI (2024), ISCAS (2024), Neurips (2023), ICCV (2023), CVPR (2023), ECCV (2022), CVPR (2022), ISCAS (2022), and GLSVLSI (2020)
- Program Committee of AAAI (2024) and KDD-2023 FL4DataMining Workshop