

Shang Wu

Google Scholar: <https://scholar.google.com/citations?hl=zh-CN&user=fBp8EWgAAAAJ>

Email: shangwu2028@u.northwestern.edu

EDUCATION

Northwestern University

09/2023-06/2028 (Expected)

- *PhD in Computer Science*

RICE University, Overall GPA:3.83/4.0

01/2022-05/2023

- *Master of Electrical and Computer Engineering*

The George Washington University, Overall GPA:3.62/4.0

09/2018-05/2021

- *Bachelor of Science in Computer Science*

PUBLICATION (Chronologically, authors with * are equally contributed)

[1] Zhongzhi Yu*, **Shang Wu***, Shunyao Zhang, Yonggan Fu, Yingyan Lin. “Hint-Aug: Drawing Hints from Vision Foundation Models towards Boosted Few-shot Parameter-Efficient ViT Tuning” In *IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR) 2023*

[2] Yang Zhao*, **Shang Wu***, Jingqun Zhang, Sixu Li, Chaojian Li, Yingyan Lin. “Instant-NeRF: Instant On-Device Neural Radiance Field Training via Algorithm-Accelerator Co-Designed Near-Memory Processing” In *60th Design Automation Conference (DAC) 2023*

[3] Yonggan Fu, Ye Yuan, **Shang Wu**, Yingyan Lin. “Robust Tickets Transfer Better: Drawing More Transferable Subnetworks in Transfer Learning” In *60th Design Automation Conference (DAC) 2023*

[4] Jyotikrishna Dass*, **Shang Wu***, Shihui Hong*, Chaojian Li, Zhifan Ye, Zhongfeng Wang, Yingyan Lin. “ViTALiTy: Unifying Low-rank and Sparse Approximation for Vision Transformer Acceleration with Linear Taylor Attention” In *29th IEEE International Symposium on High-Performance Computer Architecture (HPCA 2023)*

[5] Yang Zhao, Ziyun Li, Yonggan Fu, Yonggan Zhang, Chaojian Li, Cheng Wan, Haoran You, **Shang Wu**, Xu Ouyang, Vivek Boominathan, Ashok Veeraraghavan, Yingyan Lin. “i-FlatCam: A 253 FPS, 91.49 μ J/Frame Ultra-Compact Intelligent Lensless Camera for Real-Time and Efficient Eye Tracking in VR/AR” In *2022 IEEE Symposium on VLSI Technology and Circuits (VLSI 2022)*

[6] Yang Zhao, Yonggan Zhang, Yonggan Fu, Xu Ouyang, Cheng Wan, **Shang Wu**, Anton Banta, Mathews M. John, Allison Post, Mehdi Razavi, Joseph Cavallaro, Behnaam Aazhang, Yingyan Lin. “e-G2C: A 0.14-to-8.31 μ J/Inference NN-based Processor with Continuous On-chip Adaptation for Anomaly Detection and ECG Conversion from EGM” In *2022 IEEE Symposium on VLSI Technology and Circuits (VLSI 2022)*

[7] Haoran You, Cheng Wan, Yang Zhao, Zhongzhi Yu, Yonggan Fu, Jiayi Yuan, **Shang Wu**, Shunyao Zhang, Yonggan Zhang, Chaojian Li, Vivek Boominathan, Ashok Veeraraghavan, Ziyun Li, Yingyan Lin. “EyeCoD: Eye Tracking System Acceleration via FlatCam-based Algorithm & Accelerator Co-Design” In *49th IEEE/ACM International Symposium on Computer Architecture (ISCA 2022)*

[8] Zhongzhi Yu, Yonggan Fu, **Shang Wu**, Mengquan Li, Haoran You, Yingyan Lin. “LDP: Learnable Dynamic Precision for Efficient Deep Neural Network Training and Inference” In *tinyML Research Symposium '22*

[9] Yonggan Fu*, Shunyao Zhang*, **Shang Wu***, Cheng Wan, Yingyan Lin. “PATCH-FOOL: Are Vision Transformer Always Robust Against Adversarial Perturbations?” In *International Conference on Learning Representations (ICLR) 2022*

[10] Yonggan Fu, Qixuan Yu, Yang Zhang, **Shang Wu**, Xu Ouyang, David Daniel Cox, Yingyan Lin. “Drawing Robust Scratch Tickets: Subnetworks with Inborn Robustness Are Found within Randomly Initialized Networks” In *Neural Information Processing Systems (NeurIPS) 2021*

[11] Yutong Gao, **Shang Wu**, and Gina C. Adam. 2020. “Batch Training for Neuromorphic Systems with Device Non-Idealities.” In *International Conference on Neuromorphic Systems 2020*. ACM

[12] Cheng Zhang, **Shang Wu**, Honglu Jiang, Yawei Wang, Jiguo Yu, and Xiuzhen Cheng. “Attribute-Enhanced De-Anonymization of Online Social Networks.” In *Computational Data and Social Networks*, 256–67. Springer International Publishing, 2019

PROJECT EXPERIENCE

Robustness and model compression, Efficient and Intelligent Computing Lab, RICE
05/2021-Present

Research Assistant, Supervisor: Professor [Yingyan \(Celine\) Lin](#), RICE

- Investigated the model robustness influences on the model performance under different model sparsity by Pytorch.
- Shown that Vision Transformers are more vulnerable than Convolution neural networks by designing a patch-based attack method using Pytorch.
- Implement different pruning methods on robust/natural models and test their performance on different datasets with Pytorch to understand model robustness and sparsity performance on transfer learning.
- Solved the underfitting phenomenon on the tiny neural network by boosting tiny neural network performance with a new knowledge distillation method.
- Propose a dynamic learnable precision method by Pytorch to quantize neural networks which balance the model size and model performance.
- Optimize the Vision Transformer’s attention part using Talyor approximation and help design a hardware accelerator for the proposed structure.

Memristive neuromorphic computing, Adaptive Devices and Microsystems Lab, GWU
01/2019-05/2021

Research Assistant, Supervisor: Professor [Gina C. Adam](#), GWU

- Implemented different weight update algorithms on memristive devices by C++ and compared the performance with the simple neural networks to simulate the performance of memristive devices.
- Investigated weight update process during model training on the memristive device using C to fix the performance gap between algorithm simulation and real device.

PROFESSIONAL SKILLS

- **Software:** C/C++, Java, Python, PHP, SQL, Verilog
- **Tools and Platforms:** Linux, Git, IntelliJ, Jupyter, Slurm, MATLAB, Pytorch, Latex

Volunteering & Activities

- Reviewer, CVPR 2023

01/2023

HONORS & AWARDS

- First place award of 2022 ACM/IEEE TinyML Design Contest at ICCAD (1/150)
10/2022
- First place award of University Demonstration at Design Automation Conference 2022 (1/50)
07/2022
- First prize, Sichuan province Electronic Design Contest (Top 10% of province)
07/2017