

# RUOYU WANG

Phone: (610) 751-3050  
Email: ruw220@lehigh.edu

## EDUCATION

---

<b>PhD</b>	Lehigh University, Computer Engineering Advisor: Y. Rosa Zheng	May. 2022 - Present
<b>MS</b>	Lehigh University, Computer Engineering GPA: 3.71/4.00 Advisor: Jieming Yin, Wujie Wen	Aug. 2020 - May. 2022
<b>BS</b>	ShanghaiTech University, Computer Science GPA: 3.86/4.00 (rank 10/115) Advisor: Pingqiang Zhou	Aug. 2016 - Jun. 2020

## INTERNSHIP

---

<b>Alibaba DAMO Academy</b> Researcher in Super Resolution Group	Feb. 2021 - Jun. 2021
<ul style="list-style-type: none"><li>Implemented a DNN compression algorithm: ExpandNet. This algorithm provides an equivalent representation of a complex network. In DNN-based super resolution tasks, train a complex ExpandNet model and compress it to accelerate inference.</li><li>Optimize compressed network on TensorRT platform, then deployed the network to a real commercial product.</li></ul>	

## RESEARCH EXPERIENCE

---

<b>Underwater Media Access Control and MIMO Pre-coding</b> Supervised by Dr. Y. Rosa Zheng Underwater Acoustic Network (UAN) suffers long propagation delay. This project deploys Deep Learning (DL) models onto source nodes in the UAN. Those intelligent nodes can learn and exploit the long propagation delay. In the experiment, the co-exist network achieved normalized throughput higher than one which means there are more than one packets received per time slot. And the work can be extended to the online MIMO pre-coding.	Sep. 2022 - Present
<b>Heterogeneous Accelerator for NN Differential Privacy</b> Supervised by Dr. Jieming Yin Differential Privacy (DP) is the technique used to protect data from differential attacks. In the concept of Neural Networks (NNs), DP consists of three steps: l2 norm, clipping, and noising. Those operations introduce redundant memory movement of the per-sample gradient. In this work, a data-flow accelerator is designed and implemented by GPGPU-Sim. GPU can offload part of its work during the execution of load/store instructions. And the	Sep. 2021 – Aug. 2022

accelerator works as an additional memory space which can apply given algorithms on-the-fly. In the experiment, the GPU-to-DRAM path decreased 90.0% and 91.4% L1 and L2 cache accesses and accelerated the system by 1.35%. The project comes out a thirteen-page paper.

### **Orchestrate Scheme for Reliable NNs on Memristors**

Oct. 2020 – Sep. 2021

Supervised by Dr. Wujie Wen

Memristor-based NN accelerators may introduce large bit error rate. If using the conventional NN classification design schemes, the output maybe unreliable. This project uses an orchestrate NN design to achieve fault-tolerance. First, encode the target classes (one-shot index) to block codes, which will use the redundant bits to correct bit-flip errors. Second, the target change converted the classification problem from single-class to multi-class, the conventional loss function cross-entropy cannot train model well. The project uses focal loss which will give loss to both positive and negative labels. And the NN achieved 36% higher accuracy under the memristor noise. The project comes out a six-page paper.

## **HONORS & AWARDS**

---

Presidential Fellowship Award	Feb. 2020
Design Automation Conference System Design Contest 3rd Place Group out of 52 Groups	Jun. 2019
ShanghaiTech University Industry Practice Outstanding Team Leader	Sep. 2018
ShanghaiTech University Outstanding Student	2016, 2017, 2018
ShanghaiTech University School of Information Science and Technology Dean Scholarship	Sep. 2016

## **TEACHING EXPERIENCE**

---

### **Advanced Distributed System (ShanghaiTech, graduate-level)**

Jul. 2019 - Aug. 2019

Teaching Assistant

- Held office hours for 12 graduate students who have taken this course.
- Revised the project assignment and guided them to accomplish the distributed file system project.

### **Computer Architecture (ShanghaiTech)**

Feb. 2019 - Jun. 2019

Teaching Assistant

- Designed 2 RISC-V assignments, created problems for three exams and held office hour for students.
- Delivered a lecture about cache to 110 sophomores.

## **SKILLS LIST**

---

**Programming Languages:** C/C++, Python, Matlab, Verilog, Rust, RISC-V

**Tools:** CUDA, TensorRT, Pytorch, TensorFlow, OpenMP, MPI, Spark, CMake, LaTeX, Git