Xuhao Luo

9500 Gilman Drive, La Jolla, CA 92093
(858) 295-9996 ⋄ x3luo@eng.ucsd.edu

Education

University of California San Diego Sep 2019 - Mar 2021
M.S. in Computer Science, Department of Computer Science and Engineering
GPA: 3.88/4.0

University of Science and Technology of China (USTC) Sep 2015 - Jun 2019
B.S. in Applied Physics, School of Physical Sciences
Major in Microelectronics and Solid State Electronics
GPA: 3.69/4.3, Rank 1/11 in the major

Research Interests

Operating System, Computer Architecture, Heterogeneous Computing, Reconfigurable Computing

Research Experience

LegoMem, an FPGA-based Disaggregated Memory System Sep. 2019 - Now
Research Project at UCSD *Supervisor: Prof. Yiying Zhang*

- Working on FPGA-based distributed memory management system for system resource disaggregation.
- Implemented hash-based address translation unit for virtual-to-physical address translation on FPGA.
- Designed a reliable transport layer protocol based on go-back-N. Implemented the reliable network stack on both FPGA and host Linux server to enable reliable network communication between host and FPGA, as well as connection management for communication between multiple FPGAs and host servers.
- Implementing distributed protocol for LegoMem service such as memory migration, as well as distributed application such as in-memory K-V storage.

Design and Implementation of HLS Based Quantized Neural Network Accelerator Jan 2019 - May 2019
Graduation Project *Supervisor: Prof. Xi Jin*

- Studied the 8-bit quantization algorithm, including the quantization algorithm, the dequantization algorithm and the implementation of the 8-bit quantized convolution.
- Designed and implemented a general 8-bit quantized convolution module on Xilinx Virtex FPGA, which achieved high parallelization through array architecture, and realized memory access optimization through data reuse.
- Developed the TensorFlow C++ API for the hardware accelerator using OpenCL. Used this accelerator to accelerate the ResNet-50 CNN and achieved a speedup of 5.17x and a memory usage reduction of 66% compared with the CPU TensorFlow implementation on Xeon E5 2686.

Binary Neuron Network (BNN) Acceleration using HLS Jul 2018 - Sep 2018
Summer Internship at Cornell University *Supervisor: Prof. Zhiru Zhang*

- Designed and implemented a BNN accelerator for LeNet-5 for MNIST handwritten digits recognition.
- Applied multiple methods to improve the performance of the accelerator including parallelization, pipelining, line buffer, task-level parallelism and batch processing.
- Implemented the accelerator on Zedboard, ZC706, and AWS EC2 F1. Achieved speedups of 33x(580fps), 88x(1543fps) and 114x(2170fps) compared with the software implementation baseline on Intel Xeon 5420 CPU.

Internship

Agora.io Jun 2019 - Sep 2019
Software Engineer Intern *Shanghai, China*

- Participated in the development of CapSync, a distributed capability negotiation system for synchronizing media capability info between users. Achieved the performance of serving a maximum request of 6000 TPS.
- Developed the mechanism for broadcasting packets among different servers using C++. Used the libevent library to implement the event-driven model.
- Wrote the testing framework and tested the service using Docker Compose and Google Test.

Project

Fault-tolerant Distributed Storage System Sep. 2019 - Dec. 2019
Project for CSE224, Graduate Network System

· Implemented a cloud-based file storage system patterned on Dropbox, with a version control mechanism to achieve version consistence across different clients.
· Multiple servers are used for duplicated file storage. Achieved consistence and fault-tolerance mechanism using Raft consensus algorithm to survive server failure, datacenter failure, and network failure.

Precise Indoor Positioning for Users in a Mall Oct 2017 - Nov 2017
Alibaba TianChi Big Data Competition *Supervisor: Prof. Qi Liu*

· Determined users' position in a mall by the Wi-Fi signal strength they received on their phones.
· Analyzed more than 2 million records collected from real scenes using machine learning methods including neuron network, random forests and GBDT, with ensemble learning applied.
· Achieved an average accuracy of 89.31%, ranked 421/2845 in the preliminary round.

Skills

Tools/Framework	TensorFlow, Docker, LLVM, Google Test
Language	C/C++, Python, Rust, MATLAB, OpenCL, Verilog, HTML, JavaScript

Honors and Awards

- 2017/18 USTC Outstanding Students Scholarship, Golden Award Sep 2018
- 2016/17 USTC Outstanding Students Scholarship, Silver Award Sep 2017
- 2015/16 USTC Outstanding Students Scholarship, Bronze Award Sep 2016
- The 13^{th} Competition of Physical Research Experiment, 2^{nd} Prize Dec 2017
- The 6^{th} Aegon-Industrial Fund Scholarship Jun 2017

Leadership

- Member of USTC Student Union Sep 2016 - Sep 2017
- Branch Secretary of Chinese Communist Youth League Sep 2015 - Jun 2019