Yulun **Yao**

Computer Architecture · Machine Learning System

508 E.University Ave., Champaign, IL, 61820

(217) 200-6084 | yuluny2@illinois.edu | https://yulunyao.io

Summary

My research interest lies in the field of Computer Architecture and Machine Learning Systems.

Specifically, I am interested in building specialized systems (and hardware) for new data-intensive workloads such as Machine Learning. The system I envision will take in a specific workload as input, optimize it, and offload to a piece of specialized hardware.

There are three major (yet conflicting) research questions underlying in this flow.

1. How to build fast and energy efficient computer architecture given Post Moore's Law constraint?
2. How to make the software optimization process as effective as possible, while reducing human effort in hand-tuning?
3. How to broaden the set of input that such system can take while achieving reasonable performance in other metrics?

Although current system research community has proposed many solutions to each individual problem, they mainly solved one problem while unfortunately, imposing huge trade offs on the other two. My goal is to reduce those trade offs, and find feasible solutions for these three problems.

Education

University of Illinois at Urbana-Champaign — *Champaign, IL*

B.S. in Computer Science — *Aug. 2016 - May. 2020*

- GPA: 3.7
- Dean's List 2016-2017, 2019
- Jame's Scholar 2016-2017

Research Experience

SAMPL Lab, University of Washington — *Seattle, WA*

Undergraduate Research Assistant — *Jun. 2019 - Current*

- Advisor: Luis Ceze.
- Group link: http://sampl.cs.washington.edu/.
- Interned at UW in Summer 2019.
- Currently working on TVM Project, an automated end-to-end optimizing compiler for Deep Learning.
- Specifically, responsible for building support for sparse workloads (Graph Neural Networks) on TVM.

FPSG Group, University of Illinois at Urbana-Champaign — *Champaign, IL*

Undergraduate Research Assistant — *Dec. 2017 - Current*

- Advisor: Christopher Fletcher.
- Group link: http://cwfletcher.net/
- Worked on Morph Project, a flexible 3D-CNN accelerator.
- Worked on CPUCNN Project, a CNN inference kernel on CPU by exploiting weight repetition.
- Worked on DARPA's SDH Project, which is to create malleable hardware and software for data-intensive algorithms.
- One paper published. One paper in preparation for CPUCNN project.

Teaching Experience

Department of Computer Science, University of Illinois at Urbana-Champaign — *Champaign, IL*

CS498IOT, Internet Of Things, Course Assistant — *Sep. 2019 - Dec. 2019*

- Course Assistant for CS498IOT, Internet Of Things. Lead CA for lab 4.
- Responsible for organizing lab sections, answering questions on lab materials, and grading.

Department of Computer Science, University of Illinois at Urbana-Champaign — *Champaign, IL*

CS433, Computer Architecture, Course Assistant — *Sep. 2019 - Dec. 2019*

- Course Assistant for CS433, Computer Architecture.
- Responsible for course grading.

Industry Experience

Intelligent Medical Objects Co. *Northbrook, IL*

Software Development Intern *May. 2018 - Aug. 2018*

- Full stack software development intern, worked on both web page and database.
- Built an web application to query IMO medical terminologies.

Publication

Morph: Flexible Acceleration for 3D CNN-Based Video Understanding *Proceeding of MICRO 2018*

Kartik Hegde, Rohit Agrawal, **Yulun Yao**, Christopher W Fletcher *May. 2018*

- arXiv:1810.06807