Yulun **Yao**

COMPUTER ARCHITECTURE · MACHINE LEARNING SYSTEM

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

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

Research Statement

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?

Current solutions mainly focus on one of the problems and unfortunately, impose trade offs on the other two metrics. My goal is to reduce trade offs imposed by solutions from current research communities, 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*

- Overall GPA: **3.7**
- Dean's List 2016-2017, 2019

Publications

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

Research Experiences

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.

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 Experiences

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 Experiences

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.

Honors

2017   **James Scholar**, College of Engineering, University of Illinois at Urbana-Champaign   *Champaign, IL*

Skills

- **Programming Languages:** C++; Python; C; Ocaml; JavaScript; Verilog; R; Java; Latex;
- **Tools and Frameworks:** Tensorflow; Caffe; Git; Docker; ReactJS;

Selected Projects

TVM: End to End Deep Learning Compiler Stack   *Research Project*

Work done at SAMPL @ UW   *June 2019 to Current*

- TVM is an end-to-end deep learning compiler stack for CPUs, GPUs, and specialised accelerators. I am responsible for building support for sparse workloads (Graph Neural Networks) and adding novel optimizations on TVM. We are targeting state of the art graph libraries such as DGL.

CPUCNN: Weight Repetition Aware Kernel   *Research Project*

Work done at FPSG @ UIUC   *Jan 2019 to May 2019*

- CPUCNN project is a prototype of convolution computation kernel on CPU. We reached better performance than MKL-DNN by exploiting repeated weights and adding fine-grained optimizations. I worked on cache-load optimization and further advanced the project by investigating sparsity utilization.
- One paper in preparation.

Software Defined Hardware (SDH)   *Research Project*

Work done at FPSG @ UIUC   *Sep 2018 to Nov 2018*

- The goal of the Software Define Hardware (SDH) program is to build runtime-reconfigurable hardware and software for data-intensive algorithms. I profiled performance for several workloads, identified bottlenecks, and provided suggestions to improve.

Dead Drop: an evil chat client   *Class Project*

Work done in CS598 CLF Secure Processor Design   *Oct 2019*

- Deaddrop is a stealthy malware that setup a communication channel without tunneling through the OS, but by using side-channel timing attack (memory contention). The transmit rate is up to 12kB/sec.

USS Illini MkII   *Class Project*

Work done in CS498 RK Web Programming   *May 2019*

- A web app with some advanced front end, back end, and database functionalities to lookup statistics of an online game.

Email Assistant   *Class Project*

Work done in CS510 Advanced Information Retrieval   *Dec 2018*

- Intelligent mail assistant with politeness checking (by utilizes Stanford's politeness model), auto completion (with N-gram model).

Neural Tracker VR   *Class Project*

Work done in CS498 SL Virtual Reality   *Dec 2017*

- A VR game built in Unity3D with Oculus toolkit to enhance cognitive functions for research purposes. Supervised by UIUC Neuroscience department.