Yufan Xu

CONTACT INFORMATION

1021 University VLG,
Salt Lake City,
Utah 84108

Phone: (352) 278-6832
E-mail: yf.xu@utah.edu

EDUCATION

University of Utah, Salt lake City, Utah USA (GPA 4.0)

Ph.D Student, Computer Science, August, 2019 - Now

The Ohio State University, Columbus, Ohio USA (GPA 3.71)

Ph.D Student, Computer Science, August, 2017 - August, 2019 (Transfer to Utah)

University of Florida, Gainesville, Florida USA (GPA 3.65)

M.S., Computer Science, May, 2016

Soochow University, Suzhou, Jiangsu China (GPA 3.50)

B.Eng., Software Engineering, May, 2014

PUBLICATION

Effective Performance Modeling and Domain-Specific Compiler Optimization of CNNs for GPU
PACT 22
- Yufan Xu, Qiwei Yuan, Erik Curtis Barton, Rui Li, P. Sadayappan, Aravind Sukumaran-Rajam

Training of Deep Learning Pipelines on Memory-Constrained GPUs via Segmented Fused-Tiled Execution
CC 22
- Yufan Xu, Saurabh Raje, Atanas Rountev, Gerald Sabin, Aravind Sukumaran-Rajam, P. Sadayappan

Efficient Distributed Algorithms for Convolutional Neural Networks
SPAA 21
- Rui Li, Yufan Xu, Aravind Sukumaran-Rajam, Atanas Rountev, P. Sadayappan

Analytical characterization and design space exploration for optimization of CNNs
ASPLOS 21
- Rui Li, Yufan Xu, Aravind Sukumaran-Rajam, Atanas Rountev, P. Sadayappan

Dependence-aware, unbounded sound predictive race detection
OOPSLA 19
- Kaan Genç, Jake Roemer, Yufan Xu, Michael D. Bond

RESEARCH EXPERIENCE

Research Assistant
University of Utah **August, 2019 - Now**
- Worked on search space optimization in TVM
  Improve consistency, efficiency of TVM internal candidate configuration selection algorithm
- Worked on design space exploration for optimizing CNN for GPUs
  Prune the kennel configuration space by using data-driven analysis and design a hybrid model for configuration quick selection
- Worked on memory efficiency for large input on ML system (pytorch)
  Solve memory constraint issue of a large input image training on a single GPU
- Worked on opmin optimization pass for tensor contraction in CCSD benchmark on MLIR
  Implement a MLIR pass to reduce total number of float operation of high order tensor contraction expressions

- Worked on a tile-size optimization problem for affine programs in polyhedral model
  Build an approximate modeling method for tile size selection

Teaching & Advicing

Course Instructor
The Ohio State University **Fall, 2018, Spring, 2019**

- Instructor for two semesters of *CS1223 Introduction to Computer Programming In Java.*
- Taught the general concepts of computer programming and programming languages by providing practical experience programming in the Java.

Teaching Assistant
University of Utah **Spring, 2020**

- Teaching Assistant for *CS 6230 Parallel Computing and HPC.*
- Planned course project, graded assignments and projects.

Working Experience

Uber , Sunnyvale, CA, USA

Software Engineer(Intern) May, 2023 - August, 2023

LatentAI , Princeton, NJ, USA

Compiler Engineer(Intern) May, 2022 - August, 2022

T-CETRA , Columbus, OH, USA

Software Engineer(Intern) May, 2019 - August, 2019

Fairchild Semiconductor, Suzhou, JS, China

Application Engineer(Intern) May, 2014 - August, 2014

Service

ASPLOS '21, '22 AE committee; CGO '23 AE committee