Qiao Zhang

zhangqiaorjc@gmail.com, (347) 325-1715

Professional Experience

Software Engineer, *Tensorflow Runtime, Google Brain*   *Mountain View, July 2018 - present*
- I was a founding engineer of the New Tensorflow Runtime Project (started by Chris Lattner).
- I implemented and improved key components for Host Runtime (examples include HostBuffer, OpAttrs, eager dispatch), GPU runtime (examples include GpuResourceRegistry), Tensorflow Python and Tensorflow Serving integration (pending open source and RFCs), and C++ utilities (UniqueAny and ResourceContext
- I implemented a few TF dialect to CoreRT dialect lowering passes and the corresponding kernels. Examples include conditional control flow.
- I introduced an MLIR dialect and compiler passes for a fallback mechanism to run ops via Tensorflow Eager Runtime (pending open source).
- I worked on TPU Runtime, and designed, built and deployed a persistent compilation cache to speedup Tensorflow Serving inference server load time for TF/XLA/TPU.

Research Intern, *Systems Research Group, Microsoft Research*   *Redmond, June 2017 - Sep 2017*
- Designed, built and deployed a streaming system that can diagnose and localize failures responsible for IaaS VMs crashes using the machine learning techniques of Lasso Regression and Hypothesis Testing. Published a research paper at the top networking conference NSDI 2018.

Software Engineering Intern, *Net Systems, Facebook*   *Menlo Park, July 2016 - Oct 2016*
- Designed, built and deployed a scalable TCP incast detection system that instruments end-host TCP stack using bcc and collects TCP statistics using a streaming system.
- Measured microburst at rack switches and published a paper at IMC2017.

Software Engineering Intern, *Google*   *Mountain View*
- At MapReduce Networking team (June 2014 - Aug 2014), I designed and implemented fine-grained locking to improve transaction throughput for the 2nd generation MapReduce backend.
- At Platforms Networking team (Sept 2015 - Dec 2015), I used Integer Program Solver to synthesize network topology to achieve desired load balancing properties.

Education

University of Washington   *June 2018*
Ph.D in Computer Science and Engineering — GPA 3.9
Advisors: Tom Anderson, Arvind Krishnamurthy
Coursework: Machine Learning, Natural Language Processing, Statistical Methods, Computer Systems.
Operating Systems, Distributed Systems, Computer Architecture, Data Management
TA: Deep Learning Systems (with Tianqi Chen), Operating Systems, Computer Networks

Williams College   *June 2013*
B.A. Computer Science and Physics — GPA 3.99
summa cum laude and Phi Beta Kappa
Caltech   *Oct 2011 - June 2012*
Visiting student — GPA 4.0

GitHub

New Tensorflow Runtime: https://github.com/tensorflow/runtime
Distributed Lock Service Using Paxos: https://github.com/zhangqiaorjc/cse550

Skills

(Fluent): C++, C, Python; (Familiar): Java, C#, MySQL

Selected Publications

Deepview: Virtual Disk Failure Diagnosis and Pattern Detection for Azure. Qiao Zhang, Guo Yu, Chuanxiong Guo, Yingnong Dang, Nick Swanson, Xinsheng Yang, Randolph Yao, Murali Chintalapati, Arvind Krishnamurthy, Tom Anderson. (NSDI 2018)

High-Resolution Measurement of Data Center Microbursts". Qiao Zhang, Vincent Liu, Hongyi Zeng, Arvind Krishnamurthy. (IMC 2017).

Rack-level Congestion Control. Danyang Zhuo, Qiao Zhang, Tom Anderson, Arvind Krishnamurthy, Vincent Liu. (HotNets 2016).

One Tunnel is Enough. Simon Peter, Umar Javed, Qiao Zhang, Doug Woos, Tom Anderson, and Arvind Krishnamurthy. (SIGCOMM 2014).

A Non-invasive Tongue Machine Interface. Qiao Zhang, Shyam Gollokota, Ben Taskar, Rajesh Rao. (CHI 2014).

Awards

2008 National Mathematics Olympiad (Singapore) Silver Medal
2007 National Physics Olympiad (Singapore) Bronze Medal