Benjamin Schreiber

bjschre2@illinois.edu
(847) 707-0084

OBJECTIVE A software development or research role involving deep learning, compilers, and programming languages. Currently considering internships for Summer 2020.

EDUCATION

Master of Science, Electrical and Computer Engineering — *expected May 2021*
University of Illinois Urbana-Champaign (UIUC)
Advised by Professor Vikram Adve

Bachelor of Science, Computer Engineering — May 2019
University of Illinois Urbana-Champaign (UIUC), 3.9/4.0
Chancellor's Scholar — Minor in Physics

EXPERIENCE

Nvidia — Summer 2019
Deep Learning Intern — TensorRT team
- Converted widely-used models for speech and audio (such as BERT, WaveNet, and Tacotron 2) from PyTorch, Keras, and TensorFlow to ONNX
- Added support for ONNX LSTM nodes, enabling TensorRT to support recurrent networks converted to ONNX
- Collaborated with the maintainers of the tensorflow-onnx and keras-onnx converters to fix converter bugs
- Architected and implemented an automated workflow using Git, Docker and Python scripting to convert, validate, and benchmark a zoo of models

Apple — Summer 2018
Intern — Core Kernel team
- Developed a new application performance profiler with the DTrace framework
- Extended DTrace with new probes to allow for fine-grained thread manipulation
- Automated testing process with Python and Bash scripts to speed development
- Upstreamed 3 patches to the XNU kernel and shipped code in iOS 12

Nvidia — Summer 2017
Intern — Carmel CPU team
- Contributed to firmware that dynamically recompiles frequently executed code
- Optimized cryptographic operations and integer division
- Improved cache utilization by reducing instruction count 50% in some crypto ops
- Used Python scripts to automate directed testing process

RESEARCH

Graduate Seminar on Heterogeneous Systems — Spring 2019
Group Project Member
- Proposed usage of CPU to compute expensive neural network activations
- Collaborated with team members to architect a fast shared memory structure to enable fine-grained cooperation between accelerator and CPU
- Wrote a simulated machine learning inference workload to collect perf data

Research of Group of Professor Christopher Fletcher — 2018-2019
Undergraduate Researcher
- Analyzed security vulnerabilities in distributed machine learning systems
- Instrumented TensorFlow models and analyzed data for a proof-of-concept attack