

Syed Tousif Ahmed

<https://syed.sh/>

Email : stahmed@seas.upenn.edu

Mobile : +1-206-619-4542

EDUCATION

University of Pennsylvania

Philadelphia, PA

M.S.E. Electrical and Systems Engineering. Specialized in SoC Architecture and EDA Tools. August 2019 – December 2021

Rochester Institute of Technology

Rochester, NY

B.S. Computer Engineering. Mathematics Minor. Specialized in Deep Learning.

August 2013 – May 2018

EXPERIENCE

NVIDIA Corporation, PyTorch Core Team

Santa Clara, CA

Senior Software Engineer

April 2022 – Present

- Design and architect the core components of PyTorch's deep learning compiler for GPUs.
- Implement optimizing graph passes on high level deep learning operations.
- Design and architect optimization sets for GPU code generation.

University of Pennsylvania

Philadelphia, PA

Graduate Research Assistant

August 2019 – December 2021

- Built tools that reduce FPGA compilation time (see publications). Assisted in teaching ESE 532 SoC Architecture, and authored handouts: https://icgrp.github.io/ese532_handouts/.

NVIDIA Corporation, Deep Learning Frameworks Team

Santa Clara, CA

Deep Learning Software Engineer

June 2018 – August 2019

- Resolved several bugs and performance issues on both CPU and CUDA. Improved performance of several Random Number Generation kernels by upto 3x. A full list of contributions can be found here: <https://github.com/pytorch/pytorch/commits?author=syed-ahmed>.
- Managed CI test environment and released NVIDIA GPU Cloud PyTorch docker container monthly.
- Conducted technical interviews for software engineering positions. Participated in code reviews and collaborated with third party contributors.
- Got into top 100 PyTorch contributors list within 7 months of joining the team.

NVIDIA Corporation, Deep Learning Frameworks Team

Santa Clara, CA

Deep Learning Software Intern

May 2017 – August 2017

- Designed and implemented Universal Framework Format (UFF) Converters for TensorFlow and Caffe2, released in TensorRT 3.0 RC.
- Analyzed performance of Caffe2 kernels for Seq2Seq models and made optimizations.

NextDroid, LLC (Startup)

Cambridge, MA

Deep Learning Engineering Intern

June 2016 – August 2016

- Implemented road image segmentation models for a semi-autonomous/self-driving car.
- Created image segmentation web interface for mass data collection, that decreased data collection cost by 60%.

SKILLS

Programming Languages: C and C++, Python, ARM Assembly, Verilog.

Technologies: PyTorch, Tensorflow, CVXPY, CUDA, LLVM, Git, Docker, CMake, Xilinx Vitis, AWS, GCP.

PUBLICATIONS

Yuanlong Xiao, **Syed Tousif Ahmed**, and André DeHon. *Fast Linking of Separately-Compiled FPGA Blocks without a NoC*. In Proceedings of the IEEE International Conference on Field-Programmable Technology, December, 2020.