Yash Akhauri

Blog | GitHub | LinkedIn | Gmail | +91 78915 12802

EDUCATION

BITS PILANI | B.E. in Electronics and Instrumentation
Aug 2016 - May 2020 | RJ, India
Fluent: Python3, PyTorch, Tensorflow, Mathematica
Familiar: C++, Java, CUDA, OpenMP, Android Studio, LibGDX, Docker

EXPERIENCE

XILINX RESEARCH | Visiting Researcher
Aug 2019 - May 2020 | Dublin, Ireland

- Working on co-design of neural network topologies and reconfigurable hardware that maps to an efficient FPGA implementation without the need for a custom accelerator architecture or a scheduler.
- Developed a library for the neural network design automation, from quantized sparse training to deployment on FPGAs.
- Targeted the Jet Substructure Classification task as part of CERN LMS L1 trigger experiments, used the library to deploy models with $10\times$ lower latency than FPGA4HEP designs.
- Demonstrated quantization library Brevitas to CERN `[GitHub]`.

URANIOM | Research Intern
Jan 2019 - Jul 2019 | France (Remote)

- Implementing semantic segmentation models for face transfer across GIFs, progressive GANs for realistic UV map generation and exploring effective weight-sharing strategies for neural networks under a research collaboration.

WOLFRAM | Undergraduate Researcher
June 2018 - July 2018 | Waltham, Massachusetts

- Developed HadaNet MLPs in the Wolfram Language and worked on C OpenMP kernels for GEMM and Convolutions using the Hadamard Binarization algorithm. `[OpenMP kernel]` | `[CUDA kernel]` | `[Whitepaper]`

PAPERS & GRANTS

Towards Inference at Clock Speed: Neural Network Topologies from Hardware Building Blocks

HadaNets: Flexible Quantization Strategies for Neural Networks [Link] California | Jun 2019
Paper accepted at CVPR'19 UAVision workshop - Orals .
Delivered a "Theatre Talk" and poster at the Intel Demo Booth at CVPR'19.

Wolfram Technology Conference – Speaker Champaign, IL | Oct. 2018
Delivered a talk on my research on Hadamard Neural Networks.

Intel Nervana Early Innovators Grant $5000
Received research grant to develop Binary Precision Neural Networks and Real time Artistic Style Transfer. The technical article can be found `[here.]` The code can be found `[here.]`

Intel CVPR Travel Grant $3000

Wolfram Student Aid $2400
Received aid to attend the Wolfram Summer School and develop Hadamard Binary Neural Networks.

Intel AI Meetup – Speaker [pptx] Delhi, IN | Sept. 2018
Spoke about my research on scaling AI using Intel technologies. This event was organized by Intel.

Intel AI Academy Success Story [Link]

Intel AI DevCon – Poster San Francisco, Bangalore | May & Aug 2018
Presented posters on quantized GEMM kernels for Intel Xeon Phi