Krishna Teja Chitty-Venkata
2040 Long Rd, 165H, Ames, IA, 50010 ⋄ +1-515-203-5766
krishnat@iastate.edu ⋄ www.linkedin.com/in/kt95

SUMMARY

- Enthusiastic researcher, working at the intersection of Deep Learning and Parallel Architecture/Computing
- Current research on pruning DNNs with respect to special and general purpose hardware (CPU, GPU)
- Former Machine Learning Intern at AMD, worked on developing DNN Quantization algorithms for GPUs
- Filed a patent (Quantization Calibration Method) during my previous internship at AMD (under review)

EDUCATION

Iowa State University *Ames, Iowa, USA*
PhD in Computer Engineering. 3.5/4.0 *Aug '17 - Present*
Advisor: Dr. Arun K. Somani

University College of Engineering, Osmania University *Hyderabad, India*
Bachelor of Engineering in Electronics and Communication. 8.4/10 *Sept '13 - May '17*

ACADEMIC/PROFESSIONAL EXPERIENCE

Iowa State University *Ames, Iowa*
Graduate Research Assistant *May '18 - Present*

My research involves optimization (compression) of DNNs for efficient inference on hardware and applying to different problems (eg: fault tolerance). My research projects are as follows:

1) Designed a Pruning algorithm and a Hyper-parameter tuning method for CNN, MLP networks which are dependent on the dimensions of a systolic array. The goal of the developed techniques is to minimize the number of computation cycles of DNN forward pass on a systolic array based Neural Network accelerator (Similar to Google's TPU). Submitted the work to IPDPS '20 (Under review).

2) Studying Structural Faults and Pruning model on Faulty DNN Hardware: Examined the impact of row and column faults on a systolic array on the performance of DNNs and proposed mitigation strategies. Our results found that even one column can degrade the accuracy of the network significantly compared to a single row fault. (Presented a short version at ASAP '19; full paper submitting soon).

3) Current Project: Developing a symmetric Pruning and a uniform Quantization algorithm to compress the Neural Networks to match the SIMD nature of GPU for achieving high speedup during inference.

Advanced Micro Devices (AMD) *Austin, Texas*
Machine Learning Intern *May '19 - Aug '19*
Worked in the MIGraphX (GPU graph optimization) team to design DNN compression algorithms for enhancing performance on AMD GPUs at inference run-time. Developed quantization techniques to convert the weights of Deep Neural Networks (CNN, MLP) from floating point to integer precision for fast execution. (Publication submitting soon).

Iowa State University *Ames, Iowa*
Graduate Teaching Assistant *Aug '17 - April '18*
TA for Digital Logic Design course in fall '17 and spring '18 semesters. Responsibilities: Supervising labs, mentoring students on Verilog HDL, FPGAs, course assignments and technical projects.

Undergraduate Research Experience
1) Bachelor Thesis: Development of Embedded System based Device for Detection of Fluorine in Water.
2) Part of a team to develop signal processing algorithms to extract modulation parameters like carrier frequency, bandwidth, code rate, cycles per phase and number of phases of Low Probability of Intercept (LPI) radars.

Research Centre Imarat, Defence R&D Organization *Hyderabad, India*
Undergraduate Technical Intern *May '16 - June '16*
Worked on a project titled "Design and Simulation of Ethernet Controller on FPGA". Developed a Transmitter and Receiver of the Ethernet controller using Verilog HDL according to the IEEE 802.3 and Ethernet controller communication protocol.

Bharat Dynamics Limited, (A Govt. Of India Enterprise) *Hyderabad, India*
Undergraduate Technical Intern *Dec '15*
Under the project entitled "Study of Missile Launcher Coordinator Unit", analysed the Coordinator Unit (PCB) which is the heart of the Missile Launcher. Assembled various components of the unit and tested under varying dynamic environmental conditions.

PUBLICATION(S)

K. T. Chitty-Venkata and A. Somani, "Impact of structural faults on neural network performance," in 2019 IEEE 30th International Conference on Application-specific Systems, Architectures and Processors (Poster).

Metuku Shyamsunder, Kakarla Subbarao, Bharath Regimanu, CVSSD Krishna Teja (2017) "Estimation of modulation parameters for LPI radar using Quadrature Mirror Filter Bank," IEEE UPCON 2017 (Best paper award).

COURSE WORK (GRAD SCHOOL)

Hardware: Computer System Architecture, Applications of Parallel Computing (CS267 UC Berkeley), Design and Analysis of Algorithms, Fault Tolerant Computing, Real Time Systems, Communication Systems

Machine Learning: Probabilistic Methods, Statistics Theory for Research, Deep Machine Learning, Statistical Foundation for Machine Learning

SKILLS

Programming: C, C++, Python, Matlab
Parallel Programming: CUDA, OpenMP, working knowledge of MPI
Machine Learning Frameworks: Tensorflow, Pytorch, Keras, Scikit Learn
Deep Learning Networks/Data sets: MLP, CNN, RNN (transformers), MNIST, Cifar10, Imagenet
Other: Linux, Shell Scripting, Verilog HDL, FPGA, Gem5 and ZSim Simulators

ACADEMIC PROJECTS

Reinforcement Learning using Neural Networks: The goal of this project is to design a reinforcement learning algorithm using Q-learning with function approximation. A Multi-layer neural network is used as the Q function approximator and the environment to be trained is taken from the OpenAI Gym. The Q-learning algorithm is then compared with other algorithm (SARSA).

High Performance Cache Simulation on GPU: Developed a CPU-GPU based cache simulator and compared it with traditional CPU-only simulation. Developed the simulator using C (for CPU-only) and CUDA C (for CPU-GPU). CPU-GPU cache simulation performed better than CPU-only simulation.

Checkpoint-based Fault Mitigation Techniques for GPUs: Proposed a new fault tolerant algorithm to aid Check-pointing which attempts to reduce the communication overhead between CPU and GPU.

TEAM WORK EXPERIENCE

- Part of Dependable Computing and Networking Laboratory (DCNL) group consisting of six researchers
- Worked in a team of five people at AMD's GPU compiler developer team (MIGraphX)
- Executive Committee member at Indian Students' Association (ISU) for 2018-19 academic year

RECOMMENDATION(S)

- Dr. Arun K. Somani (Doctoral Advisor): arun@iastate.edu
- Mike Vermeulen (Manager at AMD) - LinkedIn Recommendations Section