Efficient Processing When Intelligence Moves to the Edge

Tinoosh Mohsenin
Energy Efficient High Performance Computing (EEHPC) Lab
CSEE Department, University of Maryland, Baltimore County

Friday, March 30, 2018
11:00am – 12:00pm
Showcase Room

Continuous collection and processing of vast amounts of data is becoming more common with the advancement of wearable sensors, Internet of Things (IoT) devices and cyber physical systems. Despite their potential significant impact, intelligent mobile technologies face a number of challenges for use in daily life. These devices usually stream raw data to cloud computers for computation which leads to massive storage, significant transmission power consumption, real time constraints and privacy/security issues, thus processing at the edge of sensors is becoming increasingly preferred. Processing these sensor-level data requires a variety of signal processing and machine learning tasks that come at the cost of high computational complexity and memory storage which is overwhelming for these light weight and battery constrained platforms.

In this talk I will present research solutions that enable efficient processing of machine learning tasks to improve energy efficiency and throughput without sacrificing application accuracy for wide deployment of embedded/edge processing. First, I present DeepMatter a scalable framework across algorithms, architectures and hardware to design an embedded Deep Neural Net (DNN) accelerator. DeepMatter takes any number of raw sensor data for variety of applications and classifies the events within 92% accuracy and consumes very low power. Next, I present two programmable domain-specific manycore accelerators namely PENC and BinMAC that achieve energy and speed efficiencies comparable to application-specific custom hardware through data-level and task level parallelization as well as customization of instruction-sets per-core and near-memory computing. I will show the efficiency of DeepMatter as well as PENC and BinMAC solutions for several application domains including multi-physiological processing for seizure and stress monitoring, tongue drive assistive device, air quality monitoring and vision-based situational awareness. The solutions derived at the intersection of algorithms, architectures and imeplementation allow designers to rapidly prototype and deploy the next generation of sophisticated and intelligent systems for efficient edge processing in extreme environments.

Tinoosh Mohsenin is an Assistant Professor in the Department of Computer Science and Electrical Engineering at University of Maryland Baltimore County, where she directs Energy Efficient High Performance Computing (EEHPC) Lab. She received her PhD from University of California, Davis in 2010 and M.S. degree from Rice University in 2004, both in Electrical and Computer Engineering. Prof. Mohsenin's research focus is on designing highly accurate and energy efficient embedded processors for machine learning, signal processing and knowledge extraction techniques for autonomous systems, wearable smart health monitoring, and embedded big data computing. She has over 80 peer-reviewed journal and conference publications and is the recipient of NSF CAREER award in 2017, the best paper award in the GLSVLSI conference in 2016, and the best paper honorable award in ISCAS 2017 for developing domain-specific accelerators for biomedical, deep learning and cognitive computing. She currently leads 8 research projects in her lab which are all funded by National Science Foundation (NSF), Army Research Lab (ARL), Northrop Grumman, Boeing, Nvidia and Xilinx. She has served as associate editor in IEEE Transactions on Circuits and Systems-I (TCAS-I) and IEEE Transactions on Biomedical Circuits and Systems (TBioCAS). She was the local arrangement co-chair for the 50th IEEE International Symposium on Circuits and Systems (ISCAS) in Baltimore. She has also served as technical program committee member of the IEEE International Solid-State Circuits Conference Student Research (ISSCC-SRP), IEEE Biomedical Circuits and Systems (BioCAS), IEEE International Symposium on Circuits and Systems (ISCAS), ACM Great Lakes Symposium on VLSI (GLSVLSI and IEEE International Symposium on Quality Electronic Design (ISQED) conferences. She also serves as secretary of IEEE P1890 on Error Correction Coding for Non-Volatile Memories.