

Electrical & Computer Engineering

Graduate Student Seminar

MD Golam Sarwar Murshed

Will present a talk entitled:

“Deep learning for resource-constrained settings.”

SC160

September 9, 2022

4pm

Abstract: Deep learning (DL) has achieved enormous success in a variety of application domains, such as intelligent machines, image processing, speech processing, and medical diagnostics. Efforts to deploy deep learning models on cyber-physical systems to make real-time decisions based on fast-changing environmental conditions have increased in the last few years. However, resource limitations are a key bottleneck because modern deep learning models need a lot of computational power, a large memory, and dedicated hardware to build and function at a reasonable rate. A typical solution involves offloading data to external computing systems (such as cloud servers) for further processing but this worsens latency, leads to increased communication costs, and adds to privacy concerns. To address the issue, efforts have been made to reduce the computational requirement of deep learning models and make them suitable for real-time applications in a resource-constrained setting. Such compression efforts include model quantization and pruning. Applying such methods to deep learning algorithms alleviate the above issues by reducing the number of computation required for making a decision using DL algorithms. However, quantization and pruning often have a negative impact on the accuracy of DL models. In this thesis, we first describe major research efforts where deep learning systems have been deployed in resource-constrained settings, focusing on the operational aspects including compression techniques, tools, frameworks, and hardware used in successful applications of intelligent edge systems. Then, the successful development of a resource-aware deep learning-based grocery hazard detection system using existing DL model compression techniques and the successful deployment of a compressed model in a resource-limited setting have been demonstrated. Later, we demonstrated the successful development of a deep learning-based fingerprint segmentation system that surpassed the state-of-the-art NIST NFSEG fingerprint segmenter in terms of precise segmentation of slap fingerprints and fingerprint matching. Finally, we propose methods to extend compression techniques and find better solutions for making deep learning algorithms efficient in terms of resource usage.

Bio: Md Golam Sarwar is a Ph.D. student in Computer Engineering program at Clarkson University. He holds a B.S in Computer Science and Engineering from Chittagong University of Engineering and Technology (CUET), Bangladesh, and an M.S in Electrical Engineering from Clarkson University, NY. He has professional research work experience and has worked at Samsung R&D Bangladesh for 5 years. He is the recipient of different awards from Samsung R&D for the successful design and development of different parts of IoTivity (<http://iotivity.org/>) framework. His current research is about efficient and hardware-friendly deep learning algorithms and their applications in vision, language, and speech using the edge computing paradigm.

Also on Zoom: <https://clarkson.zoom.us/j/93502963500?pwd=M0xVZGt4emp0YzZKemhHeXZHVkRmZz09>

Meeting ID: 935 0296 3500

Passcode: 534236

One tap mobile

+16468769923,,93502963500# US (New York)

+16469313860,,93502963500# US

**Co-Sponsored by IEEE student branch and HKN*

Electrical and Computer Engineering ● CLARKSON UNIVERSITY ● Potsdam, New York 13699-5720