College of Engineering and Computing
Department of Electrical and Computer Engineering
Spring 2022 ECE Distinguished Seminar Series

Robust and Efficient Deep Learning Systems at Edge

DR. "SHELLEY" XUE LIN

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
NORTHEASTERN UNIVERSITY

Brief Bio-Sketch

Dr. "Shelley" Xue Lin is an assistant professor in Department of Electrical and Computer Engineering at Northeastern University since 2017. She received her bachelor's degree from Tsinghua University in 2009 and PhD degree from University of Southern California in 2016.

Her research interests include deep learning security and hardware acceleration, machine learning and computing in cyber-physical systems, high-performance and mobile computing systems, and VLSI. Her research is supported by NSF, ONR, Lawrence Livermore National Lab, DARPA, and DOT. She got the best paper awards at HAET Workshop @ ICLR'21 and ISVLSI'14, best technical poster at NDSS'20, and 1st Places at VNN-COMP'21 and Design Contest of ISLPED'20.

FEB. 25, 11:00 AM
RSCH 163 | ZOOM

Zoom Link:
https://gmu.zoom.us/j/93244528991

Abstract

This talk presents recent work from Dr. Lin's group on deep learning security and hardware acceleration. The first part is about vulnerability of deep neural networks. The second part is to implement efficient deep learning systems at edge covering both inference and training.

The talk begins with our design of structured adversarial examples, revealing structural information through strong group sparsity and providing better interpretability of the adversarial examples. Next, I will introduce our adversarial T-shirt, the first physical world adversarial example considering deformation of non-rigid objects. This work was published as a Spotlight Paper in ECCV'20 and has been broadly featured and cited in over 100 media outlets including Communications of the ACM, The Register, Boston Globe, etc. Then, I will present our recent work in ICLR'22 on reverse-engineering of adversarial perturbations to recover the original images. From hardware perspective, deep learning systems are subject to fault injection attacks, which manipulate neural network models for misclassification. I will discuss our modeling of such attacks through ADMM (alternating direction method of multipliers).

For efficient implementation techniques of deep learning at edge, I will briefly go through our paper leveraging model compression for mobile acceleration of 3D convolutional neural networks, achieving real-time execution for the first time. Then I will introduce our series of work on deep learning quantization for FPGA using intra-layer mixed schemes and multiple precisions. The talk ends with our NeurIPS'21 Spotlight Paper about sparse training using mobile devices.