Fan Chen
Department of Electrical and Computer Engineering
Duke University

Efficient Hardware Acceleration of Unsupervised Deep Learning

Friday, February 21, 2020, 10:45 am
Location: Marvin Center 301

Abstract
Recent advances in deep learning are at the core of the latest revolution in various artificial intelligence (AI) applications including computer vision, autonomous systems, medicine, and other key aspects of human life. The current mainstream supervised learning relies heavily on the availability of labeled training data, which is often prohibitively expensive to collect and accessible to only a few industry giants. The unsupervised learning algorithm represented by Generative Adversarial Networks (GAN) is seen as an effective technique to obtain a learning representation from unlabeled data. However, the effective execution of GANs poses a major challenge to the underlying computing platform.

In this talk, I will discuss my work that devises a comprehensive full-stack solution for enabling GAN training in emerging resistive memory based main memory. A zero-free dataflow and pipelined/parallel training method is proposed to improve resource utilization and computation efficiency. I will also introduce an inference accelerator that enables developed deep learning models to run on edge devices with limited resources. Finally, I will discuss my vision of incorporating hardware acceleration for emerging compact deep learning models, large-scale decentralized training models, and other application areas.

Biography
Fan Chen is a Ph.D. candidate in the Department of Electrical and Computer Engineering, Duke University where she is advised by Professor Yiran Chen and Professor Hai “Helen” Li. Her research interests include computer architecture, emerging nonvolatile memory technologies, and hardware accelerators for Machine Learning. Fan won the Best Paper Award and the Ph.D. forum Best Poster Award at ASP-DAC 2018. She is also a recipient of the 2019 Cadence Women in Technology Scholarship.