BINGHAMTON UNIVERSITY
STATE UNIVERSITY OF NEW YORK
Computer Science

THE DEPARTMENT OF COMPUTER SCIENCE & THE COMPUTER SCIENCE GRADUATE STUDENT ORGANIZATION (GSOCS) PRESENT

INVITED SPEAKER SERIES

Dr. Jinjun Xiong
IBM

Friday, October, 16, 2020 at 1 p.m.
https://binghamton.zoom.us/my/lijunyin

A Statistical Distribution-based Deep Neuron Network Model – A New Perspective on Effective Learning

Abstract: The impressive results achieved by deep neural networks (DNNs) in various tasks, computer vision in particular, such as image recognition, object detection and image segmentation, have sparked the recent surging interests in artificial intelligence (AI) from both the industry and the academia alike. The wide adoption of DNN models in real-time applications has, however, brought up a need for more effective training of an easily parallelizable DNN model for low latency and high throughput. This is particularly challenging because of DNN's deep structures. To address this challenge, we observe that most of existing DNN models operate on deterministic numbers and process one single frame of image at a time, and may not fully utilize the temporal and contextual
correlation typically present in multiple channels of the same image or adjacent frames from a video.
Based on well-established statistical timing analysis foundations from the EDA domain, we propose a novel statistical distribution-based DNN model that extends existing DNN architectures but operates directly on correlated distributions rather than deterministic numbers. This new perspective of training DNN has resulted in surprising effects on achieving not only improved learning accuracy, but also reduced latency and increased high throughputs. Preliminary experimental results on various tasks, including 3D Cardiac Cine MRI segmentation, showed a great potential of this new type of statistical distribution-based DNN model, which warrants further investigation.

Bio: Dr. Jinjun Xiong is currently Researcher and Program Director for AI and Hybrid Clouds Systems at the IBM Thomas J. Watson Research Center. He co-founded and co-directs the IBM-Illinois Center for Cognitive Computing Systems Research (C3SR.com<http://c3sr.com/>) with Prof. Wen-mei Hwu at UIUC. His recent research interests are on across-stack AI systems research, which include AI solutions, algorithms, tooling and computer architectures. Many of his research results have been adopted in IBM's products and tools. He published more than 100 peer-reviewed papers in top AI conferences and systems conferences. His publication won six Best Paper Awards and eight Nominations for Best Paper Awards. He also won top awards from various international competitions, including the recent champions for the IEEE GraphChallenge on accelerating sparse neuron networks, and champions for the DAC'19 Systems Design Contest on designing an object detection neural network for edge FPGA and GPU devices.