CECS
CENTER FOR EMBEDDED & CYBER-PHYSICAL SYSTEMS
UNIVERSITY OF CALIFORNIA · IRVINE

CECS Seminar Series 2018

"Cost and Power Efficient Deep Neural Network Acceleration"

Jongeun Lee

Associate Professor
School of ECE
UNIST, Ulsan, Korea

Monday, February 26, 11:00am – 12:00pm

EH 2430
Host: Prof. Fadi Kurdahi

Abstract: Deep neural networks prove to be a right direction after all, but the enormous computational complexity calls for new research into novel ways of making efficient implementations of the connectionist model for small, mobile devices.
In this talk I will discuss two directions, one based on FPGA and the other based on a new computing paradigm called stochastic computing (SC). FPGA is a very capable device housing thousands of processing elements in a single chip with low power consumption. However, FPGAs are very different from GPUs, and making it easy to program FPGAs for deep neural networks is one of the key challenges facing FPGAs today. I will present a design space exploration approach to help find the best design for a given FPGA and for a specific deep neural network model.
Stochastic computing (SC) was first introduced in the 60's when computing devices were not very reliable. Unlike conventional binary number representations, stochastic computing uses a bitstream to represent a number, and is therefore inherently more error-resilient.
In addition, SC shares some good qualities with analog computing such as low power dissipation and low cost while strictly operating in the digital domain. At the same time, error fluctuation and conversion overhead are the key challenges facing SC today. I will present a new SC architecture for deep neural networks that is much more accurate and efficient than previous SC solutions.

Biography: Jongeun Lee received his B.S. and M.S. in Electrical Engineering, and his Ph.D. in electrical engineering and computer science all from Seoul National University, Korea. In 2009 he joined UNIST (Ulsan National Institute of Science and Technology), Ulsan, Korea, where he is now an Associate Professor of Electrical and Computer Engineering. Prior to joining UNIST, he worked as a postdoctoral research associate at Arizona State University, and previously worked for Samsung Electronics. His current research interests include reconfigurable architectures, compilers, stochastic computing, and deep neural networks.