IPM

سخنرانی‌های علمی

پژوهشگاه دانش‌های بنیادی
پژوهشکده علوم کامپیوتر

Hardware Support for Efficient Neural Network Acceleration

By: Dr. Mehdi Modarressi
modarressi@ut.ac.ir

Abstract

The ever-increasing demand for power-efficiency in modern semiconductor technologies, along with the rapid growth of embedded and server applications based on deep learning, has spawned new research and development activity in neuromorphic and other brain-inspired computing models, as well as in efficient hardware and software implementation of neural networks.

Major semiconductor manufacturers and designers, such as Intel, Nvidia, Xilinx, ARM, Qualcomm, and IBM have added neural network-based learning capability into their recent products. Some of these companies, including IBM, Intel, and Google, have even produced special-purpose neural network chips in recent years. There are also numerous smaller companies that design licensable IP-cores for neural network acceleration.

In this talk, the architecture of some recent neural network accelerator chips, including IBM *Synapse* and Google *TPU*, are reviewed. Then, we will introduce the CORN neural network accelerator and show how it exploits some inherent properties of neural networks (error tolerance, redundant computation, and parallelism) for low-power neural network acceleration.

زمان : پنج‌شنبه ۹۶/۵/۲۶ - ساعت ۱۵

مکان : فرمانیه - خیابان شهید لواسانی - جنب برج کوه نور- نبش خیابان فربین- پژوهشگاه دانش‌های بنیادی - طبقه همکف

***** شرکت برای عموم علاقه‌مندان آزاد است*****