Internal Faculty Seminar

Friday, April 8, 2022 (12-1pm)

ZOOM Link:

https://usc.zoom.us/j/92417517950?pwd=WUkycy90cndVQko5R3RhQ1U3STBDdz09
Meeting ID: 924 1751 7950
Passcode: 529946

Speaker:

Keith Chugg (USC)

Title:

Co-Design of Algorithms and Hardware for Deep Neural Networks

Abstract:

Neural networks are in wide use in cloud computing platforms. This includes inference and training with the latter typically performed on programmable processors with multiply-accumulate (MAC) accelerator arrays (e.g., GPUs). In many applications, it can be describable to train on an edge device or using energy efficient application specific circuits. In this talk I will present some research results on application specific hardware acceleration methods for neural networks. Pre-defined sparsity is a method to reduce the complexity of training and inference. In contrast to pruning approaches which remove edges/weights during or after training, this approach sets a pre-defined pattern of sparse connection prior to training and holds this pattern fixed during training and inference. This allows one to design the pattern of sparsity to match a specific hardware acceleration architecture. We also consider Logarithmic Number Systems (LNS) for implementation of training. With LNS, operations are performed on the log of the quantities and therefore multiplies are simplified to addition while additions are more complex in the log domain. We present some preliminary results for LNS training and highlight ongoing challenges in applying this to larger, more complex networks. In many of these approaches we borrow from the design and implementation of iterative decoders for digital communication systems.