

Seminar

[MSCE] Efficient Training Schemes for Binary Neural Networks

Convolutional neural networks (CNNs) present many computation and memory challenges for embedded devices. An effective way to reduce the complexity of a CNN is by binarizing the weights and activations of the network. This drastically decreases the size of the CNN and results in XNOR-popcount operations replacing the standard Multiply-Accumulate operations, typically performed billions of times over the course of a single inference. However, binary neural networks (BNNs) are difficult to train and can lead to a severe degradation in accuracy. To effectively train BNNs, special methods are typically applied to help gradients flow during backpropagation through the non-differentiable functions employed in the forward pass. Other researchers have completely re-invented the training approach to accommodate the complexity of BNN training.

Contact

Manoj Vemparala

Email: Manoj-Rohit.Vemparala@bmw.de

Advisors

Manoj Rohit Vemparala, Nael Yousef Abdullah Al-Fasfous