

Backpropagation with low-precision gradient accumulation

Deep neural networks (DNNs) are usually trained with gradient-based optimization techniques. While the weights are often quantized to a few bits, it is known that the gradients should be accumulated with high precision during training. This is a problem when low-precision hardware components – for example memristors – are used. In this project, we will examine the gradient accumulation needs in neural network training. We will analyse the causes of low performance when low-resolution weight accumulation is used and investigate possible remedies.

Goals & Tasks

- Review of the state-of-the-art on low-resolution weight training.
- Implementation of DNNs for standard benchmark tasks and very simple toy tasks for conceptual experiments.
- Investigating the influence of quantized gradient accumulation on the optimization.
- Implement possible strategies to improve quantized training.

Qualifications

- Interest in deep learning.
- Experience with Python based deep learning frameworks such as TensorFlow or PyTorch are beneficial.
- Registered to one of the following:
 - ✓ **Bachelor Thesis**
 - ✓ **Project, Seminar/Project**
 - ✓ **Master Thesis**

Contact

Robert Legenstein
robert.legenstein@igi.tugraz.at