Chair for Integrated Systems
TUM Department of Electrical and Computer Engineering
Technical University of Munich

Master's Thesis

Neural Network based Approximator for Fine-Grained HW-Models

Convolutional neural networks have become the state-of-the-art in many computer vision applications. These range from medical technology to robotics applications and autonomous driving. However, most modern accurate CNNs are very memory and compute intensive, particularly for edge inference platforms.

Compression of CNNs is essential for a variety of real world applications. These techniques can result in the degradation of task-related accuracy. To counteract this, compression should be reduced to an absolute minimum during normal operation. However, if bottlenecks occur during operation, e.g. due to memory limitations or excessive energy requirements, a compressed CNN must be used.

HW models are suitable for analyzing the interaction of HW platforms and CNNs. Executing these models can take up many resources on the embedded accelerator for real world application. In the context of this work a lean approximation of HW models is to be compiled for the approximation of the interaction between HW and CNN. Moreover, the selection of an appropriate CNN model suitable for resource limitations during runtime has to be accomplished.

Prerequisites

Prerequisites

To successfully complete this project, you should have the following skills and experiences:

- Good programming skills in Python and Tensorflow
- Good knowledge of convolutional neural networks, pruning and quanitization

The student is expected to be highly motivated and independent.

Contact

Nael Fasfous
Department of Electrical and Computer Engineering
Chair of Integrated Systems

Phone: +49.89.289.23858
Building: N1 (Theresienstr. 90)
Room: N2116
Email: nael.fasfous@tum.de

Advisors

Alexander Frickenstein