

Open Thesis / Project

High-performance Deep Models for Embedded Devices

Motivation

Deep networks have shown great performance in solving complex tasks in different application domains. Existing literature demonstrates that the performance of neural networks improves with increasing number of parameters, which usually results in an increased network width measured by the number of neurons in each layer. Recent research reports that increasing the network width alone by introducing sparsity and without changing the number of parameters leads to a better performance. In this project we would like to enable efficient implementation of deep networks enjoying such optimization techniques on an embedded device, such as Arduino Nano. We will start with the simplest possible neural networks to understand the basic principles of their efficient execution and how these can be realized on different hardware platforms. We recommend to have a look at the following paper to get a general idea <https://arxiv.org/pdf/2010.14495.pdf> . **Interested? Contact us for more details!**

Target Group

Students in ICE and Computer Science.

Thesis Type

Master Project / Master Thesis.

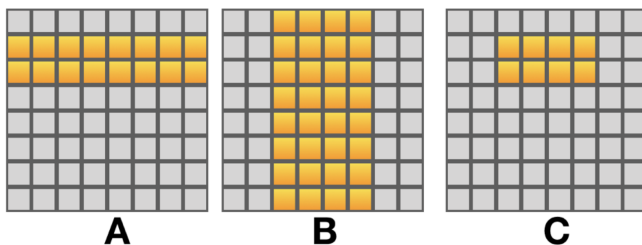


Image source: [Register blocking](#). We recommend you to have a look at the whole article.

Goals and Tasks

In this project, you will learn the impact of different optimizations on the accuracy of deep learning models and how these can be efficiently implemented on a given hardware platform. The project includes the following tasks:

- Literature review on deep learning optimizations to improve model accuracy, minimize inference time and optimize memory allocation.
- Implement a vanilla inference of a pretrained model on Arduino Nano (the platform supports TensorFlow Lite, but you will also do it from scratch).
- Test different model-specific optimizations.
- Test different hardware-specific optimization.
- Summary of the results in a written report, oral presentation.

Requirements / Skills:

- A good knowledge of embedded systems.
- Basic knowledge of deep learning or an interest to learn and understand the basics.
- Programming skills in Python and C.
- Your talent (very important!)

Contact Person

- Rahim Entezari (entezari@tugraz.at)
- Dr. Olga Saukh (saukh@tugraz.at)

