

Jan Čech  
Center for Machine Perception  
Department of Cybernetics, Faculty of Electrical Engineering  
Czech Technical University in Prague

## Bachelor's Thesis Review

Prague, August 20, 2020

**Title:** Efficient Implementation of Neural Networks for Real-Time Applications

**Author:** Matěj Suchánek

**Date received:** August 14, 2020

The thesis first presents a comprehensive review on recent techniques of efficient implementation of convolutional neural networks (CNNs). The second part of the thesis reports on experiments that were performed to measure accuracy and computational time as a function of input size. Three types of CNN problems are tested: Image classification, Object detection, Semantic segmentation. Several comparison is made among Pytorch baseline, TensorRT, TensorRT with quantization. Several hardware platforms were tested: PC CPU and GPU, NVidia Jetson CPU and GPU. Recent efficient architectures were tested (squeezeNet, mobilenet, shufflenet). The thesis concludes that for most of the problems NVidia TensorRT with quantization brings a significant speed up for a negligible loss of accuracy, while for other problems the loss of accuracy is significant.

Matěj worked on his thesis continuously and clearly acquired a competence in recent CNN implementation methods. Experiments were not trivial and required using many different softwares, running them on various platforms. Several complex test scripts were prepared. The novelty of the software tools caused some issues related to incompatibility between versions of TensorRT, CUDA, and certain CNN architectures. Nevertheless, Matěj managed to perform most of the planned experiments and presented valid quantitative results at the end. Besides coping with the technical difficulties, Matěj had to understand deeply the principles of respective methods and prepared datasets for training and calibration. I appreciate that some of the experiments were proposed by Matěj himself. Both contribution of the thesis, the review part and the experimental parts, are of a sufficient quality.

My only objection is that the progress should have been a bit faster. The text is written in a good English language, but would probably benefit from another proofreading. Nevertheless, the weaknesses are minor.

In summary, I suggest assessing the thesis by

A – excellent.

Ing. Jan Čech, Ph.D.  
Thesis Advisor