

Väinö-Waltteri Granat

Zero to DLA: Building Software Stack for Accelerating Deep Neural Networks on Custom RISC-V SoC

Abstract

Väinö-Waltteri Granat: Zero to DLA: Building Software Stack for Accelerating Deep Neural Networks on Custom RISC-V SoC

Diplomityö

Tampere University

Master's Degree Programme in Signal Processing and Machine Learning

December 2024

Lorem ipsum

Keywords: DLA, Deep-Learning, SoC, Virtual Prototype.

The originality of this thesis has been checked using the Turnitin Originality Check service.

Lyhenteet

VWW Visual Wake Words

Sisällys

1	- I/	<i>7</i>																		-
ı	n	<i>E</i> vpsyvsnayte																		
		-J F - J J J																		

1 Kypsyysnäyte

In recent years neural network based application have become more and more prominent in our everyday-life. The large driver for this has been the adoption of efficient accelerators in mobile device, that have enabled running neural network applications of mobile devices, such as smartphones.

This interest in neural networks has coincided with the industry's move to heterogeneous System-on-chip solutions being used in consumer and professional devices, to improve computational performance. More often these companies integrate their accelerators into SoCs, which include CPUs, GPUs, memory and other accelerators and peripherals in one package. Apple and Qualcomm have proved with their SoCs that they can attain desktop like performance in a smaller package than was previously possible. The industry moving towards SoCs has generated new interest in developing open-source SoCs.

The goal of this project was to build software support for the Deep-Learning Accelerator in the upcoming Headsail SoC from SoC Hub using a Renode based virtual prototype as the development platform. The goal was to use this concurrent development approach to have software support ready before the chip had been manufactured.