Summary

Author: mgr inż. Grzegorz Rafał Dec, numer albumu: d491
Supervisor: dr hab. inż. Zbigniew Hajduk, prof. PRz
Key words: FPGA, ZYNQ, LSTM, BPTT

The aim of this work is to investigate a possible implementation of RNNs, specifically LSTMs, in FPGAs, in order to achieve comparable or better calculation time to GPU or greater performance in cases where GPU or TPU cannot be utilized. A novel method was identified in this paper, presented and compared to those already in use. The approach utilized an FPGA as well as a hybrid hardware architecture, which contained a built-in processor and a programmable hardware, in a classification task of a chosen industrial process and in support of the created LSTM network's training process.

An analysis of various implementations of activation function was performed, with a focus on accuracy, calculation time and resources utilization. Subsequently, approaches already present in literature were selected for further experiments alongside the one proposed by the author. In addition to that, an LSTM cell module was implemented with multiple approaches, including novel organization of calculations, in order to examine its possible use in custom-made neural networks of any kind with an LSTM layer. In order to inspect the exhibited approach, the LSTM cell was implemented in FPGA and then tested in a real-life problem, in that case in a binary classification during failure prediction in cold forging process. Different implementation options on an FPGA were considered as well as on a hybrid hardware (software and hardware separately), specifically on Xilinx ZYNQ. During a classification task for cold forging process, in order to design and implement a digital circuit for training purposes of the LSTM network, several concepts considered in the research were further developed. All of the presented approaches were analyzed in terms of hardware availability, costs and resources required for them to operate properly. Eventually, the research was also focused on exploring the significance of accuracy of the activation functions, especially in terms of its possible effect on a classification task as well as on a training process of an LSTM network.

The results obtained in this work provide a solid basis for consideration of FPGAs and hybrid systems in LSTM network implementations, which may result in a

significant acceleration of calculations, compared even to GPUs. Also, the results have proven that it is possible to create a digital circuit in FPGA structures supporting the training of the LSTM network.