Efficient Instruction Set Architecture for Convolutional Neural Network Accelerators

Topic Description
Convolutional Neural Networks (CNNs) require highly parallel hardware structures to compute their operations efficiently. Systolic/Spatial architectures are an interesting choice for CNN execution as they do not require stochastic data movement, but rather regular, deterministic transfers and accesses [1,2]. An exploitation of this deterministic execution can be made through scheduling [3]. In most deployment scenarios, the neural networks do not change frequently, implying that a predetermined schedule would save many resources and control logic at runtime. However, to maintain the flexibility of creating such schedules for different CNNs, a coherent instruction set must be designed. The Instruction Set Architecture (ISA) will require an offline complier which determines the sequence of instructions that need to be sent out in order to execute a particular CNN.


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

- Very good programming skills in VHDL and C/C++
- Good knowledge of neural networks, particularly convolutional neural networks

The student is expected to be highly motivated and independent.

By completing this project, you will be able to:

- Understand the execution sequence of CNNs and the basic operations required by processing elements
- Test efficient ISAs that meet the required algorithmic complexity
- Evaluate the complexity of communication hardware against efficiency and latency

Contact
Nael Fasfous
Department of Electrical and Computer Engineering
Chair of Integrated Systems
Arcisstr. 21, 80333 Munich
Germany
Phone: +49.89.289.23858
Email: nael.fasfous@tum.de

This project is in cooperation with BMW AG.