# POPULAR SCIENTIFIC ABSTRACT

Nan-Sheng Huang

An Agile Design Methodology using Autobot for Hardware-based Neural Networks

Machine learning and artificial intelligence have made significant progress in many fields, becoming indispensable technologies with potential in a wide range of applications. Neural networks (NNs) are one of the most promising machine learning models. However, it is well-known that NN algorithms feature high computational complexity, being extremely computationally intensive due to the use of rich synapse connections and neuron models. For real-time embedded computing on portable edge devices, hardware accelerators can achieve low power use and fast execution times compared to Central processing units (CPUs) and graphics processing units (GPUs). However, hardware design using hardware description language (HDL) is hard due to a lower abstraction level than software programming languages. Although high-level synthesis (HLS) uses C/C++ as a design entry, the developer still needs to have solid computer architecture and logic design knowledge to harness the HLS tool. On the other hand, model retraining is unavoidable in the design iteration if there have missing use cases or uncovered corner cases of the dataset. The retraining may cause the change of the network topology and result in the modification or redesign of the hardware accelerator, which takes days to weeks with manual efforts. The change of hardware design is not as intuitive as software change.

To mitigate this issue, hardware generation is a powerful methodology on top of hardware design. Hardware generation facilitates reusability and improves design productivity by wrapping up configurable hardware design. In this work, we propose using Autobot, a light-weight software agent, to resolve real-time and low-power hardware generation problems for a class of neural networks, including MLP, ESN, and RBFNN. Firstly, scalable and configurable microarchitectures which support floating-point, half-floating point, fixed-point, and mixed-precision numbers are devised. Next, a basic Autobot is developed to parse, extract the input metadata parameters of the neural network and golden data set, configure the hardware template and test bench, and communicate with FPGA development tools for hardware generation. Furthermore, a smart Autobot with automatic holistic energy-aware design (AHEAD) methodology is developed to automatically generate the low-power fixed-point hardware accelerator, in order to eliminate the needs of developers to conduct fixed-point analysis separately.

A series of experiments in benchmark problems and proactive BMI control applications in Plan4Act are conducted to verify the effectiveness of the proposed methodology. The experimental results show that the proposed methodology can generate hardware in RTL code in less than 90 seconds for efficient and rapid design space exploration. Furthermore, the AHEAD methodology can automatically identify the fixed-point bitwidth parameters for low-power real-time hardware generation in less than 30 minutes, the saving is from days/weeks to minutes.