

Edge Computing using FPGA with the Deployment of Neural Networks for General Purpose Application

Kevini Perera¹, Chamod Hettihewa¹, Manupa Wickramasinghe^{1#}, Ashan Sandanayake¹, Chamali Rajapaksha¹, and Pubudu Pathirana²

¹Department of Electrical, Electronic and Telecommunication Engineering, General Sir John Kotelawala Defence University, Sri Lanka

²School of Engineering, Deakin University, Australia

#37-eng-0084@kdu.ac.lk

Artificial intelligence and deep learning are gaining traction in edge computing to extract insights from Internet of Things (IoT) devices. Hardware accelerators like Field Programmable Gate Arrays (FPGAs) accelerate deep learning efficiently due to their energy efficiency, parallelism, flexibility, and reconfigurability. However, resource constraints of FPGAs pose deployment challenges. This research explores hardware-accelerated applications' dynamic deployment on the Kria KV260 platform with a Xilinx Kria K26 system-on-module, equipped with a Zynq multiprocessor system-on-chip. It presents an innovative solution to dynamically reconfigure deep neural networks by running multiple neural networks and Deep Processing Units (DPU) concurrently. This research advances Edge Computing using FPGAs to facilitate efficient deployment of Neural Networks in resource-constrained edge environments.

Keywords: *FPGA, neural networks, DPU, hardware accelerator*