

ABSTRAK

Pengaplikasian Convolutional Neural Network (CNN) pada Edge Computing seringkali terhambat oleh fluktuasi latensi, tingginya konsumsi daya, dan kendala akses memori (memory wall) pada pemroses konvensional. Untuk mengatasi permasalahan tersebut, penelitian ini mengusulkan arsitektur specialized accelerator berbasis Diagonal Cyclic Array (DCA) pada SoC FPGA Kria KV260 dengan pendekatan Hardware-Software Co-Design. Prosesor ARM difungsikan menangani pra-pemrosesan citra secara dinamis dengan fleksibilitas ukuran gambar hingga 128x128 piksel di mana sistem diuji menggunakan dataset masukan MNIST beresolusi 28x28 piksel berformat 8-bit sementara Programmable Logic (PL) didedikasikan murni untuk mengeksekusi konvolusi menggunakan topologi pipelining dan line buffer sirkular guna memaksimalkan data reuse. Aliran data antar-subsistem ini dijembatani secara efisien oleh protokol AXI-Stream melalui DMA. Hasil implementasi perangkat keras menunjukkan utilisasi sumber daya yang optimal, yakni BRAM 12,15%, DSP 17,47%, dan LUT 40,76%. Sistem terbukti mampu menyelesaikan satu siklus inferensi penuh dalam 0,1115 ms (latensi perangkat keras murni 11,15 mikrodetik) dengan total konsumsi daya on-chip hanya sebesar 0,468 Watt. Hal ini membuktikan bahwa arsitektur DCA yang diusulkan sangat ideal sebagai solusi komputasi Edge AI yang sangat efisien, hemat daya, dan berlatensi deterministik.

Kata Kunci: CNN, Edge Computing, DCA, PL