ABSTRACT

In this study an optimized UNET model is used for FPGA-based inference in the context of brain tumour segmentation using the BraTS dataset. The presented model features reduced depth and fewer filters, tailored to enhance efficiency on FPGA hardware. The implementation leverages High-Level Synthesis for Machine Learning (HLS4ML) to optimize and convert a Keras-based UNET model to Hardware Description Language (HDL) in the Kintex Ultrascale (xcku085- flva1517-3-e) FPGA. Resource strategy, First in First out (FIFO) depth optimization, and precision adjustment were employed to optimize FPGA resource utilization. Resource strategy is demonstrated to be effective, with resource utilization reaching a saturation point at a 1000-reuse factor. Following FIFO optimization, significant reductions are observed, including a 55 percent decrease in Block RAM (BRAM) usage, a 43 percent reduction in Flip-Flops (FF), and a 49 percent reduction in Lookup Tables (LUT). In C/RTL co-simulation, the proposed FPGA based UNET model achieves an Intersection over Union (IoU) score of 74 percent, demonstrating comparable segmentation accuracy to the original Keras model. These findings underscore the viability of the optimized UNET model for efficient brain tumor segmentation on FPGA platforms