CSEE4840 Project Proposal

A FPGA accelerator for pose detection based on MobileNet v2

Qixiao Zhang, Tianchen Yu, Haichun Zhao, Yue Niu, Haomiao Li

Abstract:
FPGAs have gained popularity as an accelerator for CNNs due to their high-performance computing capabilities and flexibility. FPGA-based CNN accelerators can achieve high throughput and low latency, enabling real-time inference and training of large-scale models. They have shown promising results in energy efficiency and outperforming GPU-based solutions in terms of performance per watt. For our project, we plan to program the FPGA which can accelerate the computation part in MobileNet v2 to do pose detection .

Introduction:

1. MobileNet V2
   MobileNetV2 is a convolutional neural network architecture designed for mobile and embedded devices, developed by Google researchers. It is an extension of the original MobileNet architecture, which was designed for efficient inference on mobile devices with limited computational resources. MobileNetV2 improves upon its predecessor by introducing a novel architecture that uses inverted residual blocks with linear bottleneck layers, which reduces the computational cost while maintaining accuracy. In addition, MobileNetV2 employs a technique called "linear scaling of depth" to scale the model for various computational budgets. MobileNetV2 has achieved state-of-the-art performance on various computer vision tasks, including image classification, object detection, and semantic segmentation, while maintaining low latency and high energy efficiency.
2. MoveNet or PoseNet
   These are two feasible open-source implementations of doing pose detection using MobileNetV2. MoveNet is able to detect up to 17 key points of the body with low latency and high accuracy.
3. Hardware Specification
   We plan to use the FPGA part of our DE1-SOC board to accelerate the 2D convolution and pooling layers inside MobileNetV2. These layers can be arranged in pipeline order and accelerated through Verilog/C generated circuits. The HPS portion will be in charge of inter-board communication and provide output to software.
4. Software Specification

The input to the system will either come from a real-time sensor such as a Pi camera/PC camera or the images stored in the SD card on the board. The input data will be organized using the ARM core in C and then sent to the FPGA part for acceleration. After the acceleration the output results will be stored in the SD card for subsequent analysis or displayed on the VGA monitor as an additional layer on top of the original image to showcase the detected key points or pose. The end-to-end signal flow will be from SD card to SD card or from camera to monitor.

Milestone:

1. Figure out MobileNet v2 algorithm and theory of hardware acceleration
2. RTL implementation (Hardware Part) for the accelerator
3. Verification
4. Software part for user interface
5. Communication between hardware and software implementation
6. Benchmarking (possibly with Raspberry Pi)