CSEE 4840 Embedded System Project Proposal

Yangfan Wang(yw4415) Fengze Zhong(fz2393) Xincheng Yu(xy2654)

Overview

The project aims to implement a Neural Network accelerator for gesture detections. It will include development of an array of arithmetic units on FPGA for high speed inference and software codes to control the flow of the Neural Network.

CNN

We will use the Google MediaPipe Model as the starting point to understand the mechanism of the gesture recognition with CNN. However, as it is a large model containing more than 48 layers and high precision, we are still selecting a more lightweight model with pre-trained weights. This CNN flow will be conducted on the Software side as it involves high volumes of memory transactions. The computation will be offloaded to the Hardware accelerator.

Peripherals

Our design needs a USB camera to capture the image. The resolution of the image will be 640x480. We will connect this directly to the HPS on De1Soc.

Software

We will have the following functionalities from the Software side

1. Image Capturing
    a. Processes the raw data captured from the USB camera
    b. Resizes the image up to 128x128 to align with the CNN input size
    c. Pre-processes the image to adjust the brightness and contrast
2. Accelerator Control
    a. Dispatches jobs to the accelerator for starting the convolution computations
    b. Preloads the data needed for the computation to on-chip BRAM
    c. Monitors the job status of the accelerator and collect the data
3. Output result
    a. Displays the output result either to the screen or in a readable format

Hardware

On the FPGA side, we will only implement the

1. Job Controlling Unit
    a. Receives the input command from the Software side.
    b. Provides responses when the work is completed

2. Arithmetic Unit Array
    a. Includes a large matrix multiplication unit
    b. Includes a ReLu function unit
    c. Includes a natural exponential function unit
3. Memory Controller
    a. This will be connected to the Avalon Bus to receive data loading from the Software side to the BRAM.
    b. This will also be used to access the BRAM and program the registers used by the ALUs