IMPLEMENTATION OF IMAGE CLASSIFIER MODEL BASED ON CNN

Summary-

- Over the last few years there has been growing interest in exploring hardware-based solutions to solve some of the fundamental problems associated with Machine Learning techniques.
- Our focus was only to implement the trained models in hardware. This allows us to a high-level language like Python for training and directly used the trained values of parameters. The hardware implementation was done in Verilog HDL.
- Since Verilog has a module-based syntax, we started by writing modules for the simplest units (adder, multiplier, etc.) and then combined them to build larger modules.
- IEEE 754 Single Precision floating point representation was used to for data. The design can be uploaded on an FPGA which can be then used in the field for some application.
- A 3-layer of Conv2D followed by MaxPooling is used of strides (7, 7), (5, 5) and (3, 3). Then we use Flattening layer to get 64x1 vector for Dense Layer. Every CNN layer is using ReLu Activation function which is implemented as Verilog module in Fixed Point Representation of 32bits.
- Following CNN layer, 2 Deep Neural layers are added where first layer contains 8 nodes and second layer contains 2 nodes. The layer which contains 2 nodes is supposed to be output layer as number of discrete output labels must be equal to number of nodes in output layer.
- Since now a large number of multiply and addition operations are required so using the concept of pipelining would be better. There are 64 Multiplier in stage 0 which will perform the multiplication action of Matrix 8x64 to Vector64x1.
- Then 64 Adder in stage 1 followed by 32 Adder in Stage 2 and so on. Final output will be a vector of 8x1 which is to be added with biases of same dimensions. The same Multiplication and Addition procedure is followed for another layer except with different dimensions.
- For output layer, SoftMax Activation function is used which is also implemented in Fixed Point Representation of 32bits.
- Also, for layout generation, QFlow is used as the only input it needs is Verilog file with some standard cell files. Only 2 submodules layout was generated that are address_generator and sigmoid_approx.