Bryan Blakeslee
9-22-18

The goal of this project is to build a version of the LeNet5 neural network on an FPGA. Currently, the baseline Matlab behavioral model is complete and trained on the MNIST-Digits dataset. This dataset consists of a total of 70000 images of handwritten numbers ranging from zero to nine. This dataset is split into 60000 training images and 10000 testing images, in addition to the ground truth labels. Samples of the training and testing datasets are shown in Figures 1 and 2, respectively.

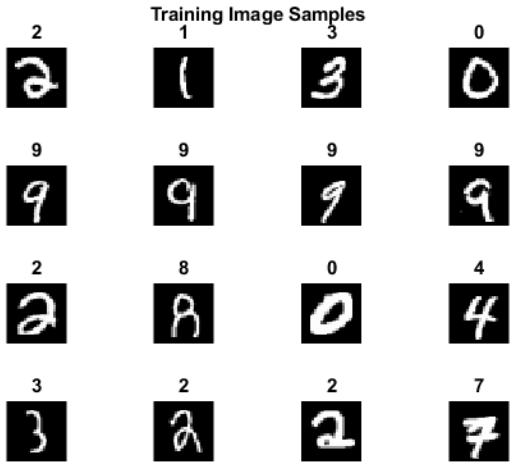

Figure 1: Sample of 16 training images with ground truth labels.

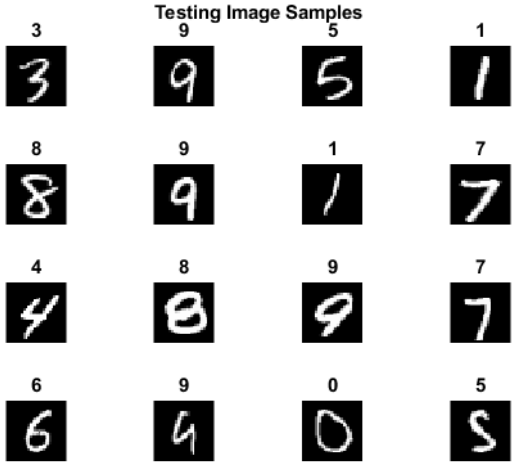

Figure 2: Sample of 16 testing images with ground truth labels.

Once these files were reprocessed into a form usable by Matlab, the LeNet5 network was trained utilizing a GTX970 GPU. Several iterations of training were performed in order to appropriately tune the network hyperparameters. The training accuracy and loss plots are shown as Figure 3.

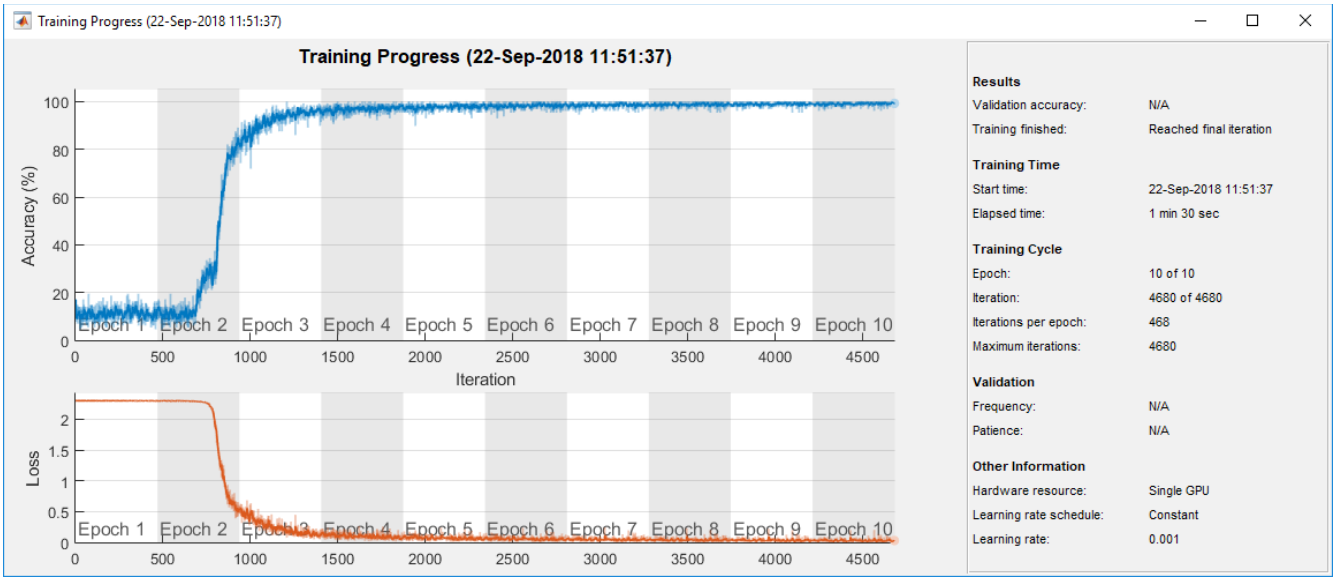

Figure 3: Accuracy (top) and loss (bottom) plots.

Once training was complete, the test set was applied to the network and the accuracy computed. The final test accuracy was found to be 98.89%. Sample output of the network classification is shown as Figures 4 and 5.

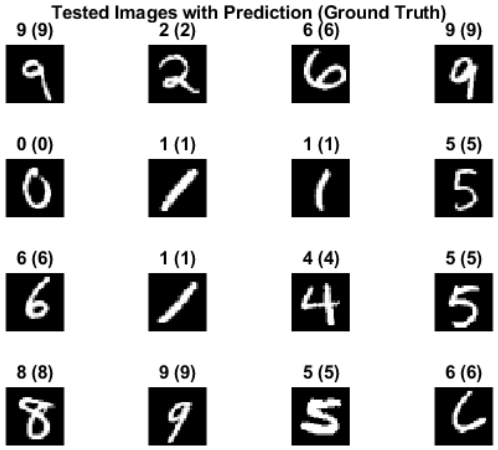

Figure 4: Randomly selected output from the network, parenthetical number is ground truth.

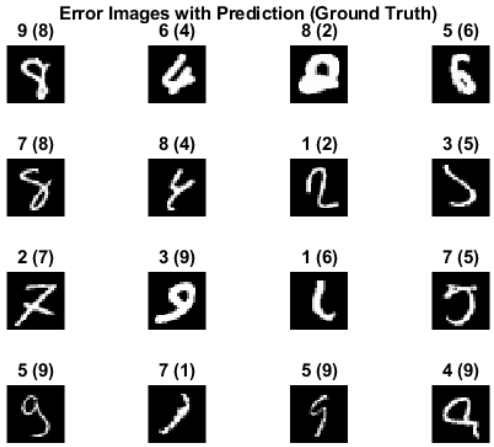

Figure 5: Randomly selected output from incorrectly classified digits, parenthetical number is ground truth.

The next step in the process is to determine how to extract the tensors from the model representation. Once extracted, these tensors will be quantized into a series of signed integer values. The network will then be retested and retrained in Matlab using these values, in order to adapt it to the modified coefficients. Once the optimal level of quantization has been determined, this particular architecture will be implemented in VHDL.

There are two significant risks currently identified. First, is that Matlab's HDL Coder tool, which allows automatic generation of VHDL code from a Matlab model, is currently not compatible with the Matlab Neural Network Toolbox, which is required to build the LeNet5 model. This will require that the network be fully implemented by hand in VHDL, which will be time intensive. Second, is that the other member of this project team has withdrawn from the course. This creates a significant amount of extra work for the remaining team member, who is now working alone.