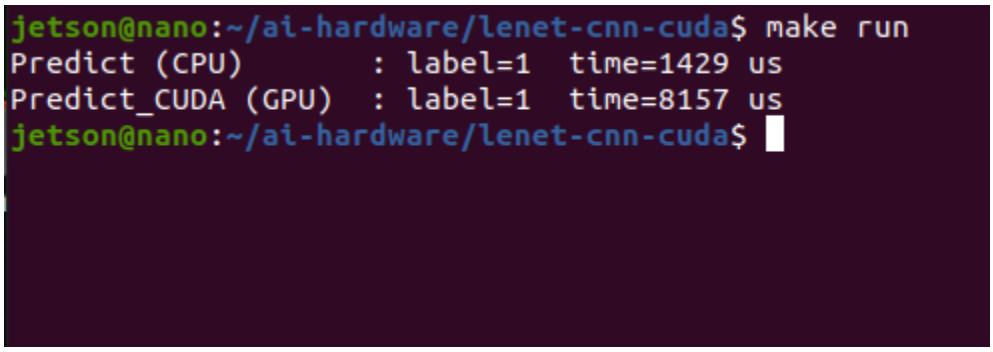
Lab 4

Group 7:

Sergio Romero, Natalia Lopez, Sebastian Castellanos, Seth Davis

The first step of the lab was to create the trained dataset using the mnist data csv file provided in the assignment details. This ended up creating 4 files. Next, was to clone the lenet repository that was also provided in the assignment details. After that, all the code and data was split up between 4 different files, build, data, model, and src. The data directory contained the datasets and the src directory contained the sequential and parallel codes. A Makefile was created to simplify the process of compiling and running all the code. Once it was verified that the sequential code would run without issues, the process of creating the parallel code began. New files cuda files called lenet.cu and lened_cuda.h were created. For lenet.cu, all of the backwards propagation was removed as it was not necessary anymore. The main focus was on forward. Several double* were created to contain the weights and biases of the dataset. After that, device memory had to be allocated for all of the data using cudaMalloc. From there, various convolutions were performed. Finally, the memory that was reserved had to be freed. The lenet_cuda.h file just contained references to a couple of functions in both lenet.c and lenet.cu. Below is a screenshot of the results. The expected results were unfortunately not met in this lab.


jetson@nano:~/ai-hardware/lenet-cnn-cuda$ make run
Predict (CPU)   : label=1 time=1429 us
Predict_CUDA (GPU)   : label=1 time=8157 us
jetson@nano:~/ai-hardware/lenet-cnn-cuda$