Part 1a:

Fully Connected Implementation:

Training Time: 1859.5388488769531

Training loss: 0.000380

Evaluation Accuracy: 45.92%

Gmacs: 1.59

Params: 1.58M

Shallow CNN:

Training Time: 2617.700516939163

Training loss: 0.002263

Evaluation Accuracy: 64.43%

Gmacs: 0.0

Params: 76.04 k

The primary difference between the fully connected network and the CNN is the number of parameters and the overall accuracy of the model. There was a nearly 20% increase in accuracy with a reduction in the number of parameters by nearly one full magnitude. The execution time for the CNN was surprisingly larger but this may just be because it is deeper than the fully connected network. I was expecting the execution time to be shorter, but it was not.

Part 1b:

Deeper CNN:

Training Time: 2682.2555429935455

Training loss: 0.435424

Evaluation Accuracy: 69.89%

Gmacs: 0.0

Params: 26.91 k

I was really surprised as to the results for this network as the training time was slightly longer, but the number of parameters was significantly reduced compared to the shallower CNN. The only thing I can think of is that the max pooling reduced the number of parameters. The other thing of note was that the loss was significantly higher, but the accuracy of the model improved by 5%.

Part 2a:

ResNet10:

Training Time: 3296.5056784152985

Training loss: 0.682870

Evaluation Accuracy: 10%

Gmacs: 0.02

Params: 272.68 k

The model has significantly more parameters than the previous CNN. The training time was significantly increased as well to represent the increase in parameters. The training loss was similar with a poor accuracy rate though. I think this is due to not having any regularization in the model leading to a very narrow learning approach.

ResNet10 Weight Decay:

Training Time: 3420.2066872119904

Training loss: 1.284183

Evaluation Accuracy: 50.52%

Gmacs: 0.02

Params: 272.68 k

The complexity of the model remained the same but had a significantly higher accuracy. This leads me to believe that there were certain parameters that were being scaled far to large and were thus throwing of the models ability to generalize.

ResNet10 WeightDecay and Dropout:

Training Time: 3473.059977054596

Training loss: 1.058628

Evaluation Accuracy: 36.15%

Gmacs: 0.02

Params: 272.68 k

We saw a reduction in the accuracy of the model when the two approaches were combined. I hypothesize this is due to too much of a reduction in the available training data.