1. Group Members:
   1. ag3743, Aman Gullati, section 001
   2. hs2923, Henrique Saboya Lopes Tavares de Melo, section 002
   3. js4637, Joaquim Lyrio, section 002
   4. rap2194, Ricardo Pommer, section 002
2. As we know, neural networks are considered to be a black box model by many people. We want to be able to understand what our neural network is doing at each layer and which information is it capturing. Thus, our problem will be to implement a feature visualization algorithm, interpret what feature map our neural network uses and how each layer reacts to the inputs.
3. We will be looking mainly at two data sets. First, we will analyse the relatively simpler MNIST data set and understand how our neural network reacts. Then, we will perform the same type of analysis for a slightly more complex dataset, CIFAR10. The MNIST data set consists of a total of 70,000 images, being 60,000 belonging to training set and 10,000 to test set. All the images are black and white (1 channel) with 28x28 pixels, hence, dimension 28x28x1. The output is a vector of size 10 since there’s a total of 10 class labels (0 through 9). The CIFAR10 data set contains 60,000 images. We plan on separating 10,000 images as test set, 10,000 as validation set and 40,000 as training set. Each input image is coloured (3 channels) consisting of 32x32 pixels, hence, with dimension 32x32x3. The output also consists 10 labels, however, this time, the labels will be animals and objects.
4. Our starting point is something that we worked on in class. The feature visualization for classification of E, F & L. We will visualize the weight values of each layer in the MNIST data set as if they were images, for a very simple neural network. This is feasible for smaller neural networks but too complex for larger neural networks. Then, our next step would be to look at the feature visualization as an optimization problem, understanding which image activates the network the most. After getting these 2 steps, we’ll be working on the CIFAR10 data, using more complex neural network structures. There are multiple methods of looking at feature visualization, at different difficulty levels, some of them are: using a deconvolutional neural network and using a regularized optimization in image space. We have been able to find some literature with detailed procedures for performing these methods.
5. The most difficult aspect of the project will be creating the feature visualization algorithm from scratch, handling the optimization of the images and the complexity of the model which may make running times quite high. We are looking at multiple papers which provide us optimization methodologies and we plan on running the code in a GPU to ensure that it runs at a good speed.