

I ended with just three layers, the initial two and then a middle, fully connected dense layer of 200. This gave me a final accuracy of around 96.958%. I tried to implement early stopping and it seems to have worked. I think there may be an issue where it's not actually stopping where it should be but it seemed to overall work. I'm thinking this since my accuracy seemed to be a bit over the place sometimes and it stopped somewhat randomly. Basically I found that the model trained a lot faster when there was more than just the two init layers in it. It went from around 94% in 40 epochs to 97% in 10 epochs.