Three Layer Network

Layer neuron counts [256, 128, 64]
Learning rate = 0.001
Batch size = 64
Optimizer = Adam
Accuracy = 96.0%

Training time
CPU times: user 11.9 s, sys: 132 ms, total: 12 s
Wall time: 12 s