Deep Learning Capstone Project

Digit Recognition in Natural Scene Images

Read Me

Source Code

This is contained in 4 files labelled from 0 to 3, which should be run in order

0_SVHNDownloadExtract.ipynb

1_SVHNExploreProcess32x32.ipynb
This is memory intensive and at times requires more than 32GB

2_SVHNTrainPredict32x3.ipynb
There are some cells in here that train on a grid so are very slow to run. You might want to skip calculating these

3_SVHNPredict32x32.ipynb

I have been running these using Jupyter Notebook which (like all the libraries below) I installed using the Anaconda package manager.

Libraries

You will need Python 2.7

My Jupyter Notebooks make use of the following libraries

- TensorFlow 0.11
- Numpy
- Scipy
- Pandas
- matplotlib
- pylab
- seaborn
- Sci-Kit Learn

I used the package manager Anaconda to install all the necessary libraries

I have put all the imports used in each sheet at the top, so you'll know before moving as soon as you run the first cell if you have anything missing.

Data

The Jupyter Notebooks use the SVHN dataset which is large, the first notebook will download the data for you to the correct folder locations. The dataset is big so the notebooks can be memory intensive and take a long time to run.