README

Yahan Zhang

7/2/2019

Description

This project converts a set of jpg and/or png images (or a mix) into mnist binary format for training (depends on imagemagick and python 2.7 PIL).

Operating System: macOS

1. Dependencies

1.1 Open Terminal, and type in the following lines to install imagemagick and the python-imaging-library (PIL):

```
brew update
brew install imagemagick php5-imagick
pip install pillow
```

2. Transform your images into an MNIST NN Ready Binary

Resize images

2.1 Select a folder and remember its path, then put **resize-script.sh** in.
2.2 Open **resize-script.sh**, and change size from "28x28" to a desirable size (e.g. 140x140). There are a total of four places.
2.3 In the same folder, create a folder named **training-images**, and another folder named **test-images**. In each of the two folders, create 10 folders named **0**, **1**, ..., **9**.
2.4 Copy-pasta your jpg and/or png images into one of the class folders, as seen in (e.g. colonial -> 0, contemporary -> 1, ... , tudor -> 9).
2.5 Put **batches.meta.txt** in the folder. Change the appropriate labels in **batches.meta.txt**.
2.6 In terminal, type in

```
sh <your path>/resize-script.sh
```

Convert images to -ubyte format

2.6 Put **convert-images-to-mnist-format.py** in the folder.
2.7 In terminal, type in

```
cd <your path>
python convert-images-to-mnist-format.py
```

2.8 Check that four files named
test-images-idx3-ubyte.gz
test-labels-idx1-ubyte.gz
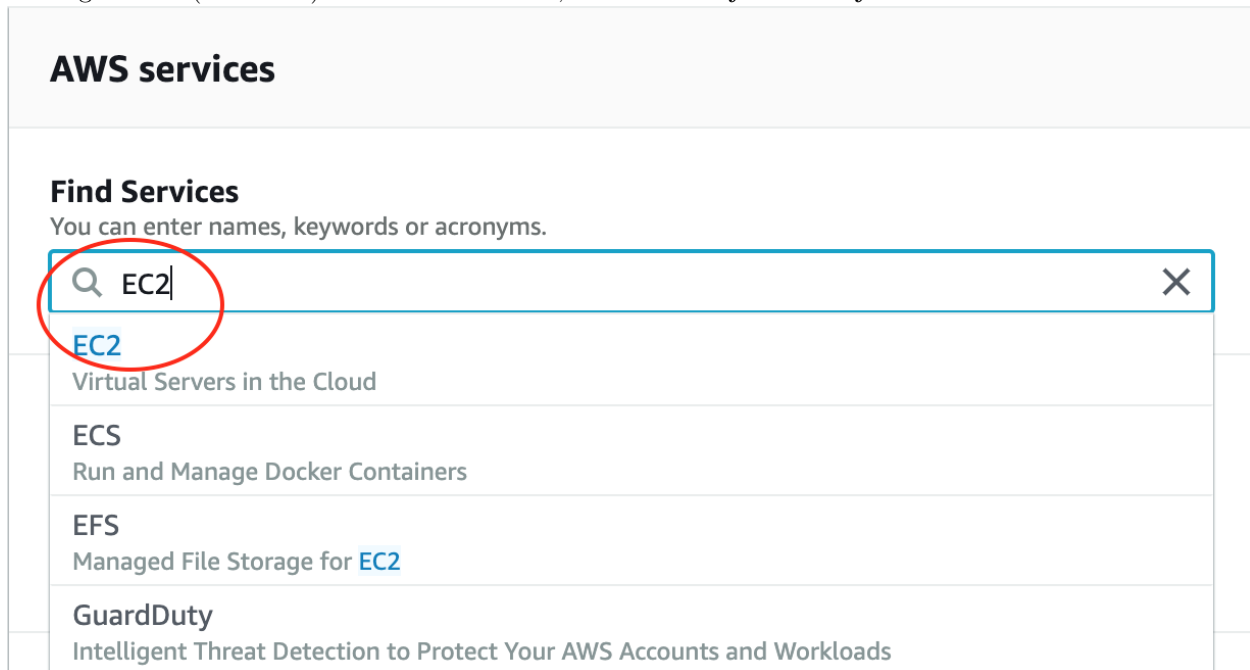**train-images-idx3-ubyte.gz**
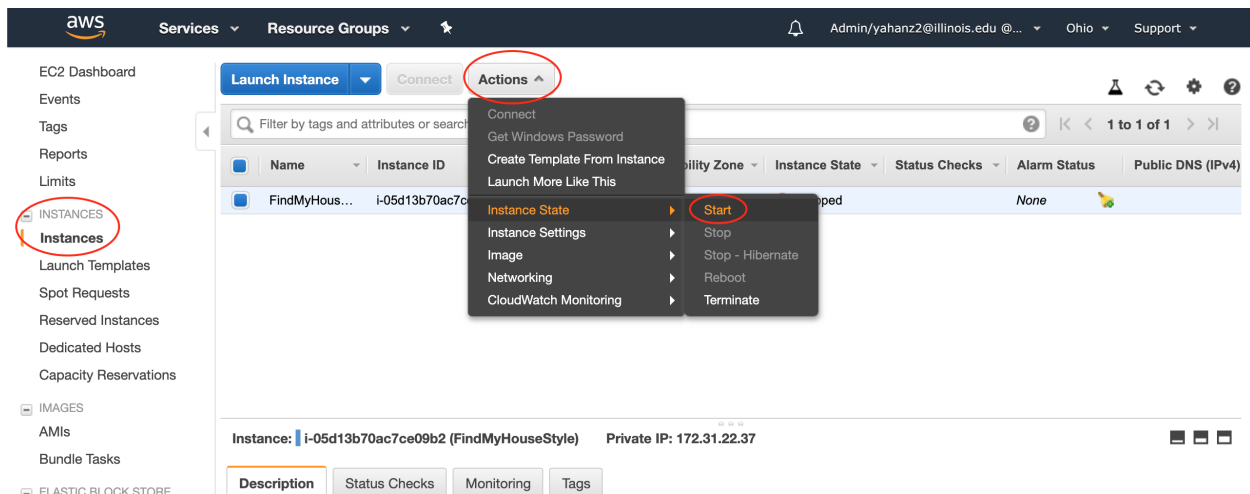**train-labels-idx1-ubyte.gz**
are in the folder.

3. Unpack images in Jupyter Notebook

3.1 Put FindMyHouseStyle.pem at a desired location, remember its path.

3.2 Login AWS (click here). In EC2 - Instances, start **FindMyHouseStyle**.

Find EC2

Launch Instance

Copy Public DNS

3.3 Under "Description" tab, copy Public DNS (IPv4).

3.4 In terminal, type in

```
cd <your path>
chmod 0400 FindMyHouseStyle.pem
ssh -L localhost:8888:localhost:8888 -i FindMyHouseStyle.pem ubuntu@<Your instance DNS>
source activate pytorch_p36
jupyter notebook
```

3.5 In terminal, look for "Copy/paste this URL into your browser when you connect for the first time, to login with a token:" and follow it.

3.6 Click "Upload" button in the topright corner. Choose the following four files in your folder and upload:

test-images-idx3-ubyte.gz
test-labels-idx1-ubyte.gz
train-images-idx3-ubyte.gz
train-labels-idx1-ubyte.gz

3.7 Make sure **Neural_Network.ipynb** and the four ubyte files are in the same folder. Open **Neural_Network.ipynb**.

Reference:

https://github.com/gskielian/JPG-PNG-to-MNIST-NN-Format