Problem: Hotel Image Classification

Training Set: 38,372 images

Test Set: 19,648 images

Machine Learning Technique used – Deep Learning

Author: Krunal Dholakia, kdholaki@uncc.edu

################################## **Required Dependency** ##################################

Need to install Theano, Keras, Open CV and Python Anaconda on Local machine and AWS EC2 Instance Local Machine.

I followed following YouTube tutorial to install Theano and Keras
**Theano**: https://www.youtube.com/watch?v=xA8Y-cfzoME
**Keras:** https://www.youtube.com/watch?v=J70j6q7aOi4

AWS EC2: GPU based Ubuntu 14.1
https://gist.github.com/graphific/f211174ebffb1f874f6d
This shell script is installing all dependency related to Deep Learning

-- Transfer this file into EC2 instance through WinSCP
-- Execute following command on AWS Instance - terminal
$chmod u+x ./<path>/<shellscript>.sh
$./<path>/<shellscript>.sh

################################## **How to replicate result** ##################################

**Step 1 -** Converted all images into 28*28 (rows*columns) size and stored them into numpy array
You can replicate this result with Load_train.py
Input: Need to provide path of train and test folder.
Output:
xdata.npy (38372, 1, 28, 28) for grey image
ydata.npy (38372, 8)
xtest.npy
imgname.npy

**Step 2 -** Transfer these array into AWS EC2 instance through WinSCP
**Step 3 -** Transfer VGG-16-Colorv1.py file into AWS EC2 instance through WinSCP

**Step 4 -** Train your network - Execute Python through following command
$python /home/ubuntu/notebook/CNNScript/VGG-16-Colorv1.py
Output:
VGG1_weights_grey111.h5 file and test.csv File.
- It will take 6 to 7 hours to train this model on GPU Based machine
Note:
This file contains other Deep Learning network also, you can select those by changing flag.
However, this file by default run for best performing model in my case.

**Step 5 -** Transfer VGG-16-Colorv2.py file into AWS EC2 instance through WinSCP

**Step 6 -** Fine tune weight - Execute Python through following command
$python /home/ubuntu/notebook/CNNScript/VGG-16-Colorv2.py
Output:
VGG1_weights_grey111.h5 file and test.csv File.
- It is to just tune weight with more iteration and avoid overfitting.
- You will get test.csv having highest score of my submission.

**Step 7 -** You can directly execute testHotelImageCLassification.py python file to get test.csv
- You need to provide path of VGG1_weights_grey111.h5 (attached in submission) file into program.

################################## **Reference** ##########################################

Nice and easy to understand theory -
http://cs231n.stanford.edu/
Keras - VGG 16 sample -
Very Deep Convolutional Networks for Large-Scale Image Recognition
K. Simonyan, A. Zisserman
arXiv:1409.1556
https://gist.github.com/baraldilorenzo/07d7802847aaad0a35d3
Special Thanks to This Guy:
https://github.com/fchollet
He made my deep learning so easy!