Ananye Patel
1 August 2018

Project AnthroPositive

Overview

Developed machine learning model to classify Naukri.com profile images uploaded by users. The model is a binary image classifier built using a Deep Convolutional Neural Network to classify an image as of a human or not. I've used Python 3.6 and Tensorflow 1.9.0 along with appropriate versions of NumPy, SciPy and Scikit-learn modules to train the model. The neural network consists of 3 convolutional layers, a flatten layer and a fully connected layer to classify an input image into either of two categories: human or not-human. The model assumes two directories in the root python directory: training_data and testing_data with appropriately labeled images. Originally the model is built to give a ratio of the likelihood of an image being a human to not a human: which is in predict.py. The network achieves a peak-classification accuracy of 98% on a test dataset.

Details

Model built using Tensorflow 1.9.0 Low Level API. OpenCV used for image preprocessing.
Other modules used:

Numpy 1.15.0	(for passing images as numpy arrays & general math functions)
Scikit-Learn	(used shuffle utility to shuffle image data consistently)
Tensorboard	(used to visualize learning)

Dataset

Size of Train-set (**Balanced)**:
Total Positives (human images) = 3491
Total Negatives (not human images) = 3395

Size of Test-set:
Total Positives (human images) = 1380
Total Negatives (not human images) = 300

Accuracy = 98.1% of images correctly identified by model
Precision = 99.5% of positively identified images (human) that are actually correct
Recall = 98.2% of actual positives (human images) identified correctly
F1 score = 98.8% : weighted average of precision and recall

Confusion Matrix:

n = 1680	Predicted: NOT HUMAN	Predicted: HUMAN	Total:
Actual: NOT HUMAN	TN = 294	FP = 6	300
Actual: HUMAN	FN = 25	TP = 1355	1380
Total:	319	1361	
