Team Info

CMPE-258 Fall 2022 Team Project - Face Recognition

- John Monsod (#015234505)
- Qiong Wu (#014640938)
- Xialu Zou (#011353316)
- Jie Liu (#015331121)

Description

Objective #1:

(Notebook #2) Given a person's image, identify who they are, based on previously trained images of people. Implement using Raw CNN model training to predict for a multi-class classifier of faces. Then, leverage a pre-trained FaceNet model with Nearest Neighbors to predict on newly supplied faces (our team member's faces).

Objective #2:

(Notebook #1) Given two separate person images, determine if they are the same person or different persons. Leverage a pre-trained FaceNet model with MTCNN to generate face array of features for a subset of LFW matched/unmatched pairs to verify if two images are of the same person or different people.

Dataset

Labelled Faces in the Wild (LFW)

LFW is a large collection of celebrity faces useful for training & testing 13,233 total images, 5749 unique persons, 1680 persons with multiple images. Within our implementation, we used a subset of the dataset due to limited computational resources.

How to run the notebook #1 for evaluating the Facenet on LFW

Notebook #1 Filename: cmpe258-teamproject-notebook1.ipynb

1. We recommend running the notebook in **Google Colab**.
2. Before running the code, please install MTCNN python package, the command provided in the first block.

```
!pip install mtcnn
```

3. Please authorize the Colab to connect with a Google drive.

```
from google.colab import drive
drive.mount('/content/drive')
```

4. Set the "basepath" value to your project base directory in the google drive.

```
basepath = '/content/drive/My Drive/cmpe258/lfw'
#This is an example, please substiute it with your own path
```

5. Put the pre-trained Facenet model weights into "{basepath}/model", and the notebook will load the model from this path.
6. Run the notebook.
7. The step "Use MTCNN face detector and Facenet embedding to get distance from 400 image pairs" may take hours. Please be patient and keep the kernel alive until it finished.

How to run the notebook #2

Notebook #2 Filename: cmpe258-teamproject-notebook2.ipynb

For both parts described below, we recommend running it in **Google Colab**.

Pre-requisites

- Create a Google drive directory where LFW dataset will be stored, i.e. `/content/drive/My Drive/lfw`.
- Upload the LFW dataset into that designated `lfw` directory.
- Create a separate `working` directory where processed data will be stored, i.e. `/content/drive/My Drive/working`.
- Create `face_dataset` and `face_query_set` directories for the FaceNet portion of this notebook.
- Upload the corresponding GitHub images into the `face_dataset` and `face_query_set` directories.

Part 1 - Raw CNN

Follow the execution of each step according to Part 1. It can be seen with the following main steps: * Data exploration - dataset distribution across persons, number of unique images, face features using MTCNN, etc. * Data preparation - preparing the top 6 persons in LFW for a multi-class classifier into working sub-directories. * CNN implementation - defining convolution, max-pooling and flattening layers into neural network with multi-class softmax outputs. * Validation and performance evaluation - running through multiple epochs of training, validation of test data, and evaluation of overall performance.

Part 2 - FaceNet with Nearest Neighbors

Follow the execution of each step according to Part 2. It can be seen with the following main steps: * Load the pre-trained FaceNet model as specified.

* Build the face database consisting of embedding and label pairs, using the `face_dataset` directory containing additional images to supply FaceNet. * Display the original images in `face_dataset` to verify visually how the images look like. * Run the 3 separate prediction tests. Each is supplied with a specific image in the `face_query_set` directory, and a prediction is made based on the previously submitted embedding and labels, which applies Nearest Neighbor algorithm to do so.