A Project Presentation On Applications Of Deep Learning In Biometric Recognition:Facial Recognition

Objective

- Real time multiple face recognition from video, applying the concept of One shot learning using Siamese network and Triplet Loss.

Abstract

Digitalization is not a future-deemed technology anymore. More and more manual jobs are being increasingly replaced by electronic/digital mediums. Such scenarios lead to an increased demand for a better approach towards fast, secure, reliable, and accurate user identification and authentication to name a few.

In this project, we have shown one such example where we use facial recognition to act as a substitute. We used deep learning techniques for the fulfillment of the end product. The resulting system is fast and accurate, thus aiding those applications which require face recognition.

Introduction

What is Biometrics?

Biometric can be defined as distinct,significant and measurable characteristics of a human being that can help in identification,verification and recognitipon of a particular individual’s identity.

Biometrics measurements include :

Physiological:

- Finger-scan
- Facial Recognition
- Iris-scan
- Retina-scan
- Hand-scan

Behavioral:

- Voice-scan
- Signature-scan
- Keystroke-scan

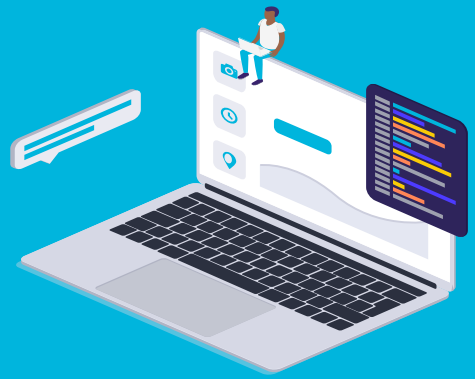
Why Facial Recognition?

Face recognition is a method of identifying and verifying the identity of the individual using one’s face.
This system uses biometrics to map the facial features from an input(image/video) to find a match in the database with known faces.
The reasons to chose facial recognition:

- It requires no physical interaction with the user. It is of quite an importance specially in a crisis like that of COVID-19 pandemic, etc.
- Facial recognition cannot be forged.
- It is accurate and allows for high enrolment and verification rates.
- It does not require an expert to interpret the results.

FRAMEWORKS AND LIBRARIES USED

- Tensorflow
- Numpy
- Matplotlib
- Keras
- Dlib
- OpenCv



Deep Learning

Deep Learning is a sub-field of Artificial Intelligence that uses artificial neural networks to learn features from the input data. The word 'deep' signifies the many hidden layers used in the neural network architecture.

What is Neural Networks?

A neural network can be defined as a series of algorithms that attempt to acknowledge fundamental associations in a set of data. The neural network can be said to be functioning like a ‘neuron’, it can adapt to changes without the need to redesign the whole training set.

One-Shot Learning

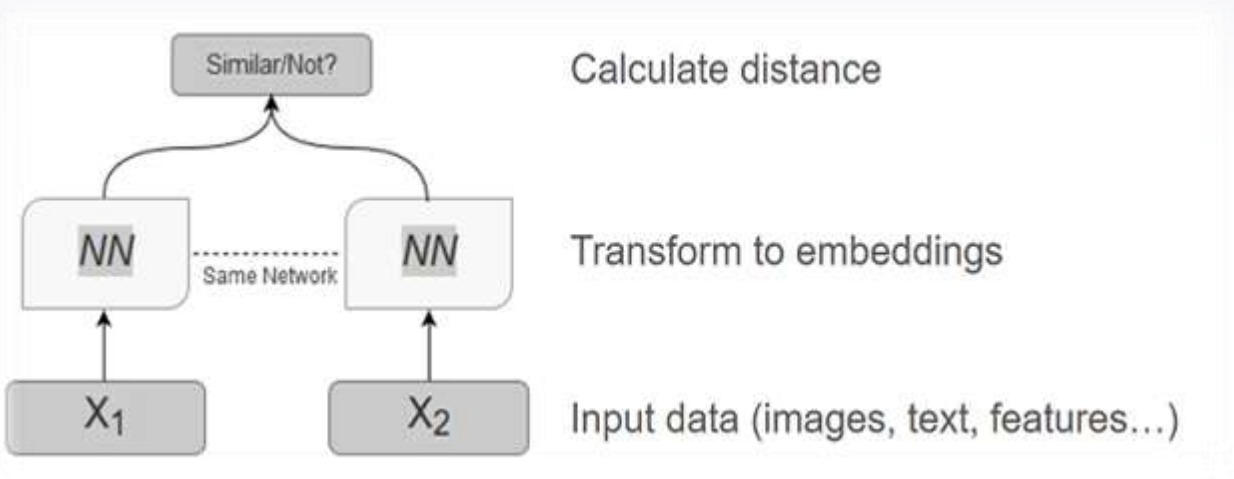
One of the drawbacks of a traditional deep learning algorithm is its requirement of a large dataset for tasks like image classification and recognition. Such a large dataset cannot be gathered if we want to train a DL network to identify a person. Also to include a new class that needs to be classified, we will need to retrain the entire model.

In situations like a large office that may see new hiring and layoff regularly, it will not be possible to employ a decent DL based model due to the above limitations.

One-shot learning aims to rectify these drawbacks. The main idea behind one-shot learning is to train a network to identify certain characteristics from an image instead of that exact image and use these pre learned characteristics to classify new unseen data. Its advantage being, once the network has learned these 'characteristics' then it requires only a single image to include a new class, and also it is not required to train the entire model to include this new class.

Siamese Network

A deep learning architecture that works significantly well with one-shot learning is the siamese network. It is a neural network that contains multiple instances of the same model and share the same architecture and weights..



Standard Siamese Network

The basic idea here is to take an input,extract its embedding by passing it through a neural network(CNN) and continue the process with different inputs.Then compare the embeddings to check whether there is a similarity between them.

Given an input image xi,the neural network outputs this into 128 dimensional encoding f of xi. So we learn the parameters so that if two pictures, xi and xj, are of the same person, then the distance between their encodings will be small .And, if xi and xj are of different persons, then the distance between their encodings will be large.

If $x^{(i)}, x^{(j)}$ are the same person, $\|f(x^{(i)}) - f(x^{(j)})\|^2$ is small.

If $x^{(i)}, x^{(j)}$ are different persons, $\|f(x^{(i)}) - f(x^{(j)})\|^2$ is large.

Facenet

FaceNet is the name of the facial recognition system that was proposed by Google Researchers in 2015 in the paper titled *FaceNet: A Unified Embedding for Face Recognition and Clustering*. It is proposed as an approach in which it generates a high-quality face mapping from the images using deep learning architectures such as ZF-Net and Inception. Then it uses a method called triplet loss as a loss function to train this architecture.

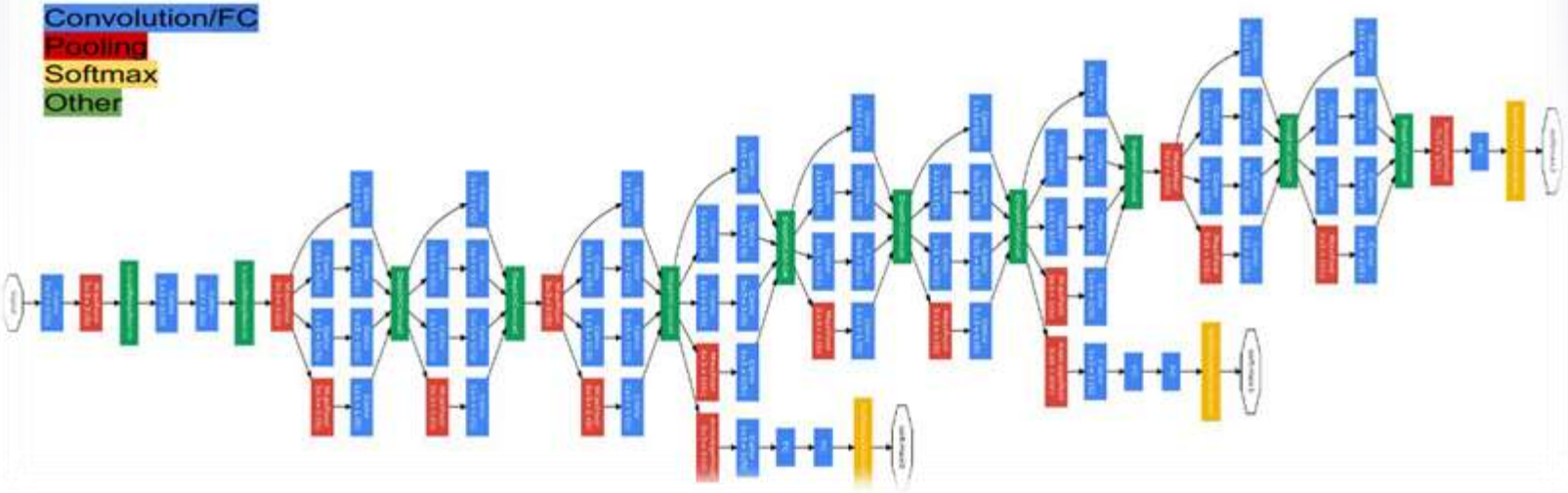
Openface

OpenFace is the open source implementation of Facenet.

Model

- Facenet is trained using a siamese network consisting of 3 inception networks. These networks are trained simultaneously and their weights are shared among each other. The output of each network is a 128D encoding of the input image. This encoding is then passed to the triplet loss function which calculates the loss and updates the learning parameters for the network.

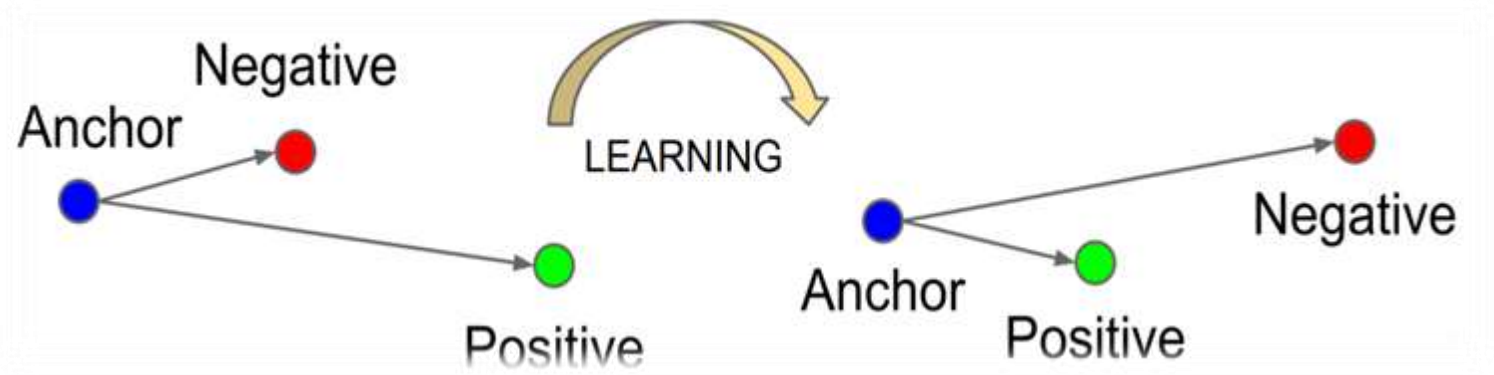
Full Inception-v1



Triplet Loss

Triplet loss is a loss function for machine learning algorithms where a baseline (anchor) input is compared to a positive (truthy) input and a negative (falsy) input. The distance from the baseline (anchor) input to the positive (truthy) input is minimized, and the distance from the baseline (anchor) input to the negative (falsy) input is maximized.

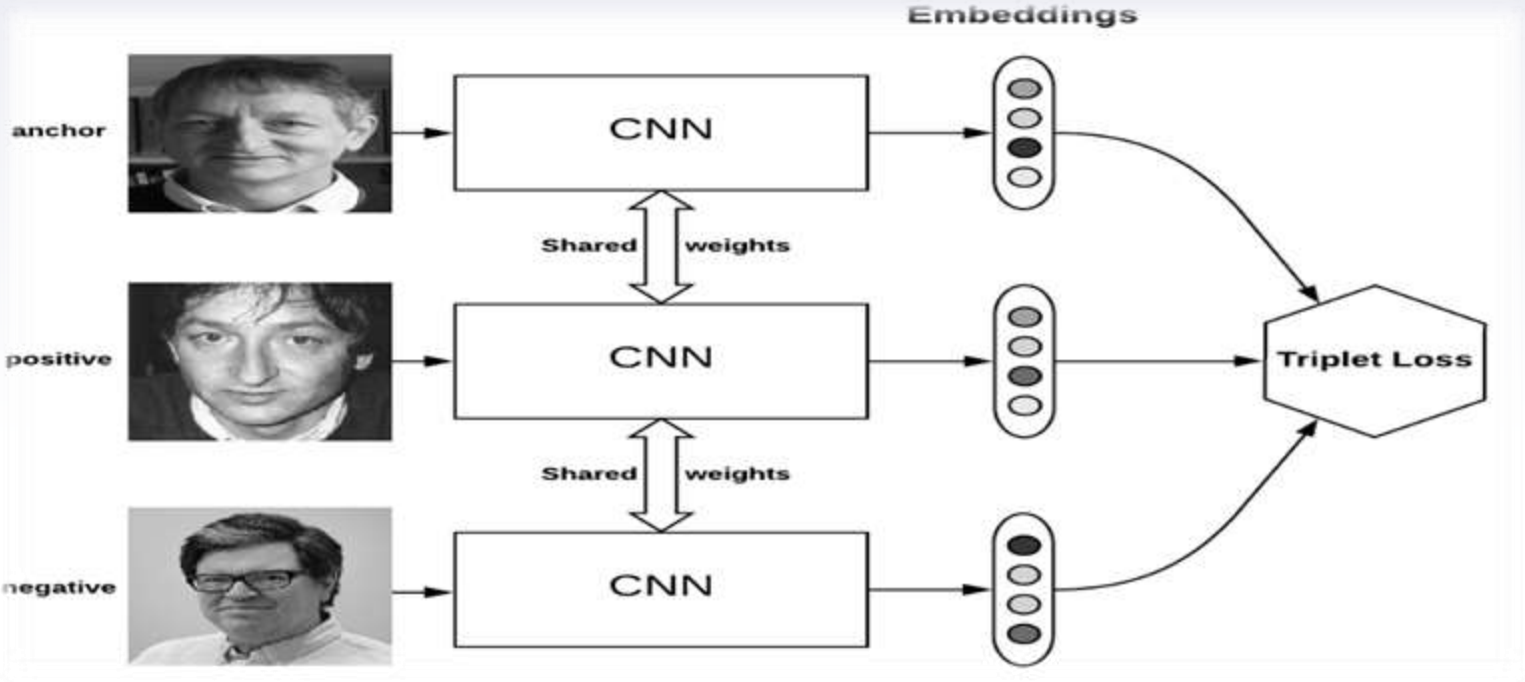
.



Considering the task of training a neural network to recognize faces, the network is trained to output a distance which is small if the image belongs to a known person and large if the image belongs to an unknown person. However, if we want to output the closest images to a given image, we would like to learn a ranking and not just a similarity. A triplet loss is used in this case.

$$\sum_{i}^{N} \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]$$

Now, the model can learn to make the same encoding for different images, which means that distances will be zero, and unfortunately, it will satisfy the triplet loss function. For this reason, we can add a margin alpha (hyperparameter), to prevent this from happening, and to always have a gap between A and P versus A and N.



Triplet loss on two positive classes and one negative class

Datasets

- The original Facenet model is trained on about 260M images
- OpenFace is trained using about 600,000 images combined from the large datasets CASIA-WebFace [YLLL14] and FaceScrub [NW14]
- In our implementation we use the openface model

Methodology

In our implementation, we use transfer learning on the open-source model of Facnet I.e Open face. We load our model and select the images of the persons we want to save in our database. We then generate encoding for these images and save these encodings in our database instead of the images. Each person corresponds to a unique encoding stored in the database.

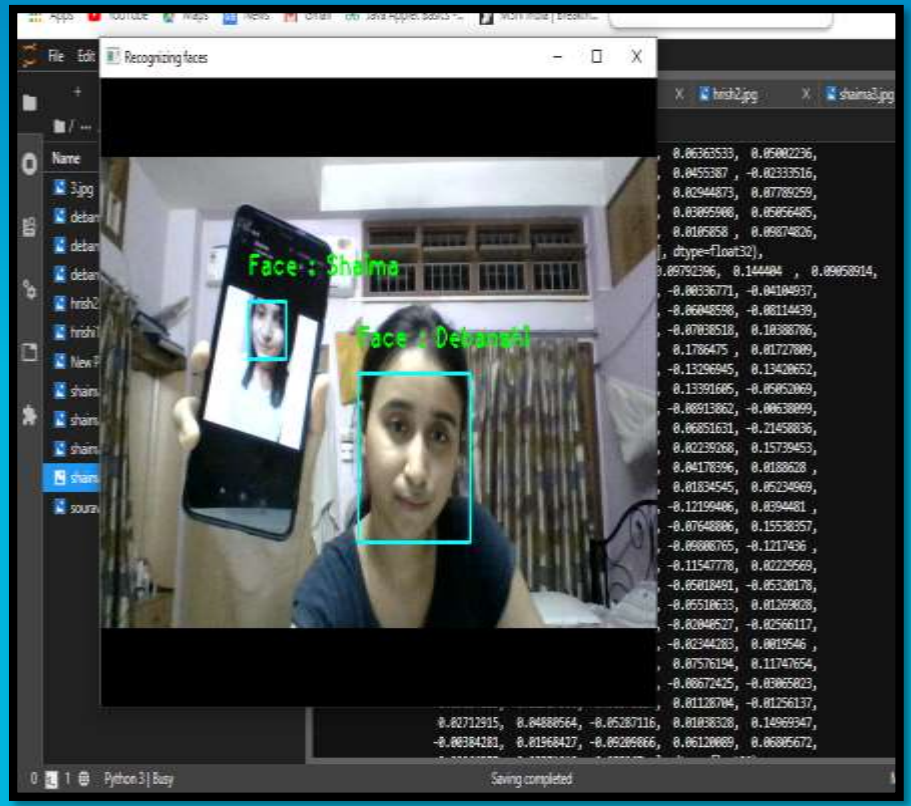
When detecting live faces we capture each frame from the camera and run it through dlib to detect and align the faces present in the frame. An encoding is generated for each face and this encoding is compared with all the encodings stored in the database. If the generated encoding is within a certain specified threshold matches with an encoding stored in our database, we return the encoding label which contains the person's name. Thus we can recognize this person.

Outputs and Results

```
0.66745174 it's Shaima
0.12501578 it's Hrishi
0.7344601 it's Debanshi
[47]: <matplotlib.image.AxesImage at 0x7fad9409edc0>
```

Face detection from still image

Outputs and Results



Face detection from live video

Future Scope

Considering the circumstances where PINs, signatures can be forged or guessed easily despite repeated warnings and cautions fuel to the already existing security breach.

This heightens the need to provide for a system that guarantees to prove the identity of the person accurately.

Applying this facial recognition system we want to provide simple and secure alternatives to pin or other ID-based recognition systems.
Also, develop an application that can be used in the fields such as attendance management, security, etc that guarantees high accuracy along with its current feature to detect faces from live video.

Conclusion

Our project aimed to develop a facial recognition system that can detect faces accurately using one-shot learning technology. To develop the system, we were able to learn the basics of deep learning and its application in biometrics.

Successfully executing this model on the facescrub and casia-webface dataset, we can detect any faces from live video feed with the given accuracy.

References

1. FaceNet: A Unified Embedding for Face Recognition and Clustering

https://arxiv.org/abs/1503.03832

2.OpenFace: https://github.com/cmusatyalab/openface