IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁒ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.11480

Face recognition using Siamese neural networks by one shot learning

Adarsh Goswami[1], Abi Dogra[2], Anubhav Bajpai[3], Anish Tripathi[4]

Inderprastha Engineering College, AKTU[1-4]

Abstract: Face recognition is a classical problem in computer vision. With the recent outbreak of covid-19 across the globe, there is much focus on face recognition systems as contact biometric methods are unsafe. To implement a face recognition system, the model must be robust and precise. In this paper, we review the past studies of face recognition. Second, this paper implements a one shot learning facial attendance system. Various network architectures are explored to improve the accuracy.

1. INTRODUCTION:

A facial recognition system is a technology capable of matching a human face from a digital image or a video frame against a database of faces, typically employed to authenticate users through ID verification services, works by pinpointing and measuring facial features from a given image. With the recent outbreak of Covid-19, face recognition models are being emphasized as these systems are contactless compared to contact biometrics methods like fingerprint scanner etc. Until the 1990s facial recognition systems were developed primarily by using photographic portraits of human faces. Research on face recognition to reliably locate a face in an image that contains other objects gained traction in the early 1990s with the principle component analysis (PCA). The PCA method of face detection is also known as Eigenface and was developed by Matthew Turk and Alex Pentland [1]. Eigenfaces are determined based on global and orthogonal features in human faces. A human face is calculated as a weighted combination of a number of Eigenfaces. Because few Eigenfaces were used to encode human faces of a given population, Turk and Pentland's PCA face detection method greatly reduced the amount of data that had to be processed to detect a face. In 1997 the PCA Eigenface method of face recognition [2] was improved upon using linear discriminant analysis (LDA) to produce Fisher faces. Real-time face detection in video footage became possible in 2001 with the Viola–Jones object detection framework for faces [3]. Paul Viola and Michael Jones combined their face detection method with the Haar-like feature approach to object recognition in digital images to launch AdaBoost, the first real-time frontal-view face detector [4].

The methods for face recognition kept updating but the necessity of huge data remained constant. In this study, we aim to build a robust One shot learning face recognition attendance system. This approach can overcome the need of huge datasets as it works by taking a single training image for each individual person.

2. RELATED STUDIES:

One shot learning may be defined as a method in which only a single training instance is required for each individual in the dataset. This is achieved by training a siamese neural network which outputs an embedding vector for each image. The network is trained so as to minimize the distance between embeddings of similar faces and maximize the distance between embeddings of dissimilar faces.

Face Net

Face Net[5] provides a unified embedding for face recognition, verification and clustering tasks. It maps each face image into a euclidean space such that the distances in that space correspond to face similarity, i.e. an image of person A will be placed closer to all the other images of person A as compared to images of any other person present in the dataset.

This optimization problem can be defined in two ways :

Triplet Loss:

In this method, we take 3 images in an instance. The first image is called anchor image, second is a positive example of anchor image(i.e. Image of same person in anchor image) and third is negative example of anchor image(i.e Image of different person that person in anchor image).

DOI: 10.17148/IJARCCE.2022.11480

The Loss function is defined as

$\| f(A)-f(P) \|^2 - \| f(A)- f(N) \|^2 + \alpha \leq 0$

Where, f(A) is embedding of anchor image f(P) is embedding of positive image f(N) is embedding of negative image α is margin factor

The margin factor helps us to create a lower bound for the distance between the Anchor-Positive and Anchor-Negative pair.

Binary Loss:

Each training instance contains 2 images and a label, if the images are of the same person, the instance is labeled as 1 and 0 if vice versa. Both the images are passed through the same network. The logistics unit at the end gets the aggregate of feature vector of both the images:

$| f(X(i))- f(X(j)) |$

Where, f(X(i)) is embedding vector of first image f(Y(i)) is embedding vector of second image

The logistics unit treats the aggregate of feature vectors as it’s Regression parameter Output is given as

$G(w^*|f(X(i))- f(Y(i))| + b)$

3. METHODOLOGY

FaceNet can be implemented in various architectures as its primary goal is to output an embedding vector irrespective of the network architecture.

Various faceNet architectures and their performance is given in table(1)

Architecture	Validation
Zeiler & Fergus (220×220)	87.9%
Inception (224×224)	89.4%
Inception (160×160)	88.3%
Inception (96×96)	82.0%
mini Inception (165×165)	82.4%
tiny Inception (140×116)	51.9%

We chose Inception(160x160) for the sake of balance in Validation accuracy and processing speed of the model.

The euclidean embedding output of the model is fed to a Cosine distance function which yields the distance score between two image instances.

The Cosine distance function is defined as

$$1 - \frac{\sum_{i=1}^{n} A_i \, . \, B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \, . \, \sqrt{\sum_{i=1}^{n} B_i^2}}$$

Where, A is embedding vector of first image

B is embedding vector of second image

Images can be classified similar or dissimilar based upon the threshold value of cosine distance function defined by the user.

For masked images, the images are passed to Dlib library to detect key mask points of the face. Next, the apt and chosen mask image is picked and warped according to the key face points. The image is masked and stored into database. The next image which comes for testing is checked in both masked and unmasked database.

4. RESULTS AND DISCUSSIONS

The dimensions of the embedding also affect the performance of the model, we tried the given dimensions in table(2)

Dimension of embedding	Validation
64	86%

128	88%
256	87%
512	85%

We chose an embedding dimension of 128 as it has the highest validation accuracy and is mediocre in time wise performance.

The value for cosine distance threshold is totally based upon the use case, a sensitive system will have low distance threshold and will improve the model's precision score whereas in a well generalizing model, the distance threshold should be comparatively high to increase recall score of the model. The scores with various distance thresholds are given in table(3)

Distance threshold	Accuracy	Precision	Recall
0.2	70.65%	0.982	0.703
0.5	80.7%	0.804	0.796
0.6	82.35%	0.793	0.815
0.7	73.45%	0.701	0.853

For general purpose usage, threshold of 0.6 is optimal as it has mediocre values for both Precision and Recall.

5. CONCLUSIONS AND FUTURE DIRECTIONS

Face recognition is a classical problem in Computer Vision that still remains to be fully discovered. In future, the face recognition systems would have power to recognize a person in harsh environments or even when some parts of the images are occluded.

6. REFERENCES

[1] Malay K. Kundu; Sushmita Mitra; Debasis Mazumdar; Sankar K. Pal, eds. (2012). Perception and Machine Intelligence: First Indo-Japan Conference, PerMIn 2012, Kolkata, India, January 12–13, 2011, Proceedings. Springer Science & Business Media. p. 29. ISBN 9783642273865.

[2] Jun Wang; Laiwan Chan; DeLiang Wang, eds. (2012). Neural Information Processing: 13th International Conference, ICONIP 2006, Hong Kong, China, October 3-6, 2006, Proceedings, Part II. Springer Science & Business Media. p. 198. ISBN 9783540464822.

[3] Malay K. Kundu; Sushmita Mitra; Debasis Mazumdar; Sankar K. Pal, eds. (2012). Perception and Machine Intelligence: First Indo-Japan Conference, PerMIn 2012, Kolkata, India, January 12–13, 2011, Proceedings. Springer Science & Business Media. p. 29. ISBN 9783642273865.

[4] Li, Stan Z.; Jain, Anil K. (2005). Handbook of Face Recognition. Springer Science & Business Media. pp. 14–15. ISBN 9780387405957.

[5]F. Schroff, D. Kalenichenko and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 815-823, doi: 10.1109/CVPR.2015.7298682.