Human Verification Using One Shot Learning

Nikhil Verma
Departpem Of Information Technology
National Institute Of Technology Karnataka
Surathkal, Mangalore, Karnataka, India
nikhilverma.222it026@nitk.edu.in

Abstract—**Humans are the only species that can learn from a single instance, and one-shot learning algorithms aim to replicate this unique skill. On the other hand, performance often depends on having a large number of annotated training examples per class, notwithstanding the great performance of Deep Learning-based approaches on numerous picture classification challenges. The use of deep neural network-based systems in many practical applications, such as facial recognition, is undoubtedly hampered by this problem. Additionally, the system will need to be completely retrained in order to accommodate the inclusion of a new class. The strength of deep learnt traits, however, also could not be disregarded. In this study, the ideal deep learning features will be combined with a conventional one-shot learning framework for verification of person using their facial image.**

Keywords—Embedding, Siamese Neural Network, Facenet.

I. INTRODUCTION

A matched matching is used in face verification or authentication to compare a provided face image to a reference face image whose identity is being asserted. One-to-many matching is used in face identification or recognition to compare a given face image to all the reference face images contained in the data and determine the identity of the given face image.

Over the course of the last few decades, face recognition has been thoroughly investigated. Its importance in a wide range of other digital applications, including security, digital entertainment systems, video analytics for marketing, and video indexing from streaming videos, cannot be overlooked. Earlier face recognition was primarily based on hand-crafted features like SIFT (scale invariant feature transformer), SURF (speeded up robust features), Local Binary Pattern, Histogram of Gradient, and Fisher vectors, like any other image analysis problem. However, with the introduction of deep-learning methodologies, there has been a shift towards Deep Neural Networks. In those early days, research was concentrated on enhancing the pre-processing stage, adding local descriptors, and feature transformation, but these methods fell short of addressing the difficulties of unrestricted face recognition. Handmade feature-based techniques were employed. deep learning Using a cascade of processing units for feature extraction and transformation enables the learning of various levels of representations and abstractions. this leads to various representation of information about the image. in the emotion, lighting, and position of the face. One drawback of deep-learning methods is the requirement for a large amount of annotated data while learning to recognize a face. when a new class needs to be added to model, it requires large amount of annotated pictures of that class to retrain the deep learning model to recognize images of new class. Although transfer learning strategies assist in reducing such issues by freezing the first few layers adjusting weights that have already been trained from the last few layers on new images, it does not entirely solve the issue of large amount of training data requirement. Few-shot or one-shot algorithms are designed to do classification requiring a few samples of the images. Consequently, a creative combination of transfer learning and one-shot learning provide a detailed feature representation created with deep learning methods and supplying a One-Shot learning framework with those representation for categorization. a widely used approach to employ one-shot learning techniques is using a Siamese neural network and Triplet loss function.

In this project we will be using a face recognition model trained on large amount of facial data for providing embedding of the images and we will use triplet loss function for creating a database of users by registering with their images. Later when a person's image is given for verification we will search the database for the said person and provide result in form of how much the person is similar to a person in database. I will be applying the one shot learning technique for recognizing the person with face mask.

This paper is organized in the following order: Introduction(sect. I), Literature Survey(sect .II), Proposed Methodology(sect. III), Result & Analysis(sect. IV), Conclusion(sect. V).

II. LITERATURE SURVEY

For nearly three decades, approaches for face detection and recognition have been a prominent focus of image analysis research. An article, which was published in the early 1990s, the authors use a limited number of 2-D Eigenvectors to represent faces. Face recognition techniques can be roughly categorized as handmade features-based approaches and subsequently deep-learning technologies-based approaches. The hand-crafted methods largely concentrated on the extraction and reduction of high-dimensional artificial characteristics. Principal Component Analysis, Linear Discriminant Analysis, and manifold learning techniques like Locality Preserving Projection are examples of representative dimension reduction techniques. The representative method following the development of deep learning was to directly learn the discriminative face representations from the original image. For instance, Hu et al's introduction of the convolutional neural network used for facial recognition is an excellent example. It examines the benefits and drawbacks of this approach and lays out the future development path. this work is further investigated, and cutting-edge results are obtained. Even while CNN performs exceptionally well in some applications, many real-world applications that demand learning from or drawing conclusions from small quantities of data, class imbalance, and adjusting to a continuous intake of new class information are difficult for such algorithms to handle. The challenge of scaling up an effective, reliable face recognition system is also not an exception in this situation.

There have been a number of works that tackle this issue in recent years. Guo et al .'s innovative underrepresented classes promotion loss term, which aligned the norms of the weight vectors of underrepresented classes and normal classes to give the one-shot classes an equivalent weight-age, was proposed as a solution to the data imbalance problem. In their study, Wang et al. offer a system based on CNN that addresses the problem of insufficient training data by applying a balancing regularizer, shifting the center regeneration, and adjusting the clustering center. However, the network performs badly because to a lack of training data and an imbalance in the data. Ding et al .'s solution to the underrepresented class issue in one-shot learning focused on using generative models to generate additional cases. By modifying the data variances and adding features from other normal classes, it presented a generative model to synthesize data for one-shot classes. The method of deep attribute encoding of faces for one-shot face recognition was proposed in another paper by Jhadav et al.. They honed a deep CNN for face recognition using specific features of human faces, such as the face's shape, hair, and gender. Their experimental findings using common datasets demonstrated improved performance of deep attribute representations. in the situation of two one-off facial recognition methods like a one-shot similarity kernel and exemplary SVM. Wu et al suggested utilizing a hybrid classifier framework The nearest neighbour (NN) model and CNN. The author's Using a domain adaptation network, Hong et al. the writers produced photographs in order to answer the One-shot task To train the deep model, different poses are used with a 3D face model. An imposed softmax with these components was proposed by Zhao et al. L2 normalization, optimum dropout, selective attenuation, and model-level improvement that increased the baseline softmax function to create a more accurate depiction of few-shot learning.

Bromley et al. suggested the idea of Siamese Networks for the signature verification problem, and demonstrated the application of deep convolutional Siamese networks for one-shot tasks with a notable level of accuracy. Face detection, feature extraction, and recognition are the typical steps in face recognition. I'll create a feature vector to represent each face using an architecture based on FaceNet. In this study, we offer a way for creating a reliable face verification system that makes use of the deep learnt feature characteristics by combining the idea of Deep Convolutional Siamese Networks and a transfer learning strategy.

III. Dataset for face Recognition model

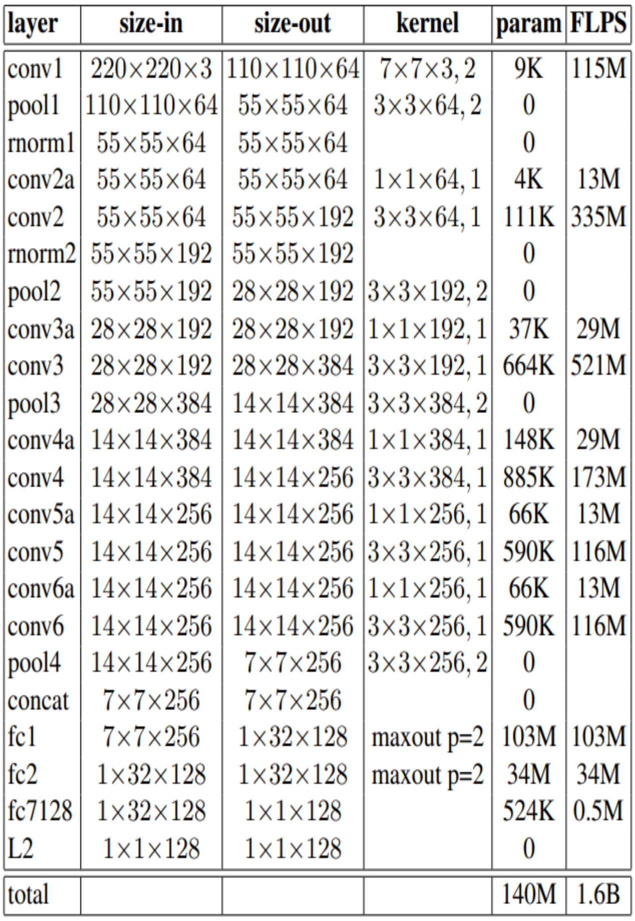
I am using LFW (labeled faces in the wild) dataset which consist of several images of 5749 peoples for training the facial recognition model. The LFW data set is provided by University Of Massachusetts, Amherst. The LFW dataset is considered as public benchmark for face verification.

IV. Proposed Methodology

There are various methods for achieving one-shot learning. In this study, we investigated two strategies: Siamese neural network-based strategy first one is a deep feature encoding approach and second is classification of the encoded features based on their nearest neighbours. We chose a strategy by fusing the two methods. The Improvised technique uses a combined method that uses embedding generated using inception model as input siamese network. in the database identity of person and their feature vector will be stored feature vector is generated by inception model. and when a person comes for verification his/her image taken and converted into feature vector using the same inception model. and the feature vector will be compared with the feature vectors already in the database.

A. Training Facial Recognition Model –

I am using LFW labelled image data to train the FaceNet model. FaceNet is a facial recognition system that was proposed by Google Researchers in 2015. The structure of the model is in the following table.



layer	size-in	size-out	kernel	param	FLPS
conv1	220×220×3	110×110×64	7×7×3, 2	9K	115M
pool1	110×110×64	55×55×64	3×3×64, 2	0	
rnorm1	55×55×64	55×55×64		0	
conv2a	55×55×64	55×55×64	1×1×64, 1	4K	13M
conv2	55×55×64	55×55×192	3×3×64, 1	111K	335M
rnorm2	55×55×192	55×55×192		0	
pool2	55×55×192	28×28×192	3×3×192, 2	0	
conv3a	28×28×192	28×28×192	1×1×192, 1	37K	29M
conv3	28×28×192	28×28×384	3×3×192, 1	664K	521M
pool3	28×28×384	14×14×384	3×3×384, 2	0	
conv4a	14×14×384	14×14×384	1×1×384, 1	148K	29M
conv4	14×14×384	14×14×256	3×3×384, 1	885K	173M
conv5a	14×14×256	14×14×256	1×1×256, 1	66K	13M
conv5	14×14×256	14×14×256	3×3×256, 1	590K	116M
conv6a	14×14×256	14×14×256	1×1×256, 1	66K	13M
conv6	14×14×256	14×14×256	3×3×256, 1	590K	116M
pool4	14×14×256	7×7×256	3×3×256, 2	0	
concat	7×7×256	7×7×256		0	
fc1	7×7×256	1×32×128	maxout p=2	103M	103M
fc2	1×32×128	1×32×128	maxout p=2	34M	34M
fc7128	1×32×128	1×1×128		524K	0.5M
L2	1×1×128	1×1×128		0	
total				140M	1.6B

The layers conv 1, conv 2a, conv 2, conv 3a, conv 3 conv 4a, conv 4, conv 5a, conv 5, conv 6a, conv 6 fc1, fc2 are inception blocks.

B. Triplet Loss

Triplet loss function is used as loss function

For an image xx, we denote its encoding $f(x)$f(x), where ff is the function computed by the neural network.

Training will use triplets of images (A,P,N):

- A is an "Anchor" image--a picture of a person.
- P is a "Positive" image--a picture of the same person as the Anchor image.
- N is a "Negative" image--a picture of a different person than the Anchor image.

These triplets are picked from our training dataset. We will write $(A_{(i)},P_{(i)},N_{(i)})$ to denote the i-th training example.

$$\mathcal{J} = \sum_{i=1}^{m} \Big[\underbrace{\| f(A^{(i)}) - f(P^{(i)}) \|_2^2}_{(1)} - \underbrace{\| f(A^{(i)}) - f(N^{(i)}) \|_2^2}_{(2)} + \alpha \Big]_+$$

Here, we are using the notation "$[z]$+" to denote $max(z,0)$.

C. Image to encoding –

The trained FaceNet model is used to extract embedding of the input image. The embedding of the input image is a 128 dimensional vector. For registering a person we map their embedding to their name and store in a database.

D. Verification –

When a person comes for verification his/her image is taken and an embedding of the image is computed using the same FaceNet model used in above step. And the distance between the embedding is measured and based on a threshold it is decided whether the person is who he/she is claiming to be.

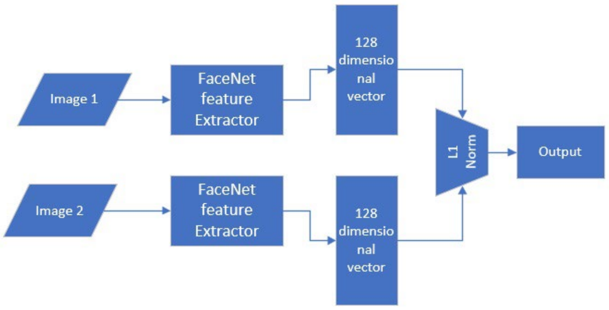
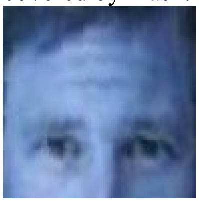


Fig. Siamese Network Model

E. Verification on masked faces

In an image with face mask the face mask is not a feature for recognition only the part of face above the mask can be used as features. So by cropping image we will generate new picture with only the part which are not covered by mask.



For verification as there is less features on the image the threshold value need to decreased for accurate prediction.

V. Result and Analysis

For testing the verification system I am using 30 images of different persons and during experiments it was found that for unmasked images the threshold value of 0.7 is giving best performance and for masked images the threshold value of 0.5 is giving best performance.

VI. Conclusion

In this project I have implemented a strategy to verify and recognize a person with only single image required for training. I have also given solution to use the same pretrained model to verify and recognize masked images. This model can be applied in scenarios like building entry gate verification, attendance system etc. as there are various model for face recognition which have been trained extensively using large dataset and in powerful computers we can use those models generating feature vector and use those feature vector for verification and recognition tasks using the Siamese neural network architecture.

References

[1] Koch, Gregory, Richard Zemel, and Ruslan Salakhutdinov. "Siamese neural networks for one-shot image recognition." *ICML deep learning workshop*. Vol. 2. 2015.

[2] 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.

[3] Y. Taigman, M. Yang, M. Ranzato and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1701-1708, doi: 10.1109/CVPR.2014.220.

[4] Zhao, Wenyi, et al. "Discriminant analysis of principal components for face recognition." *Face Recognition*. Springer, Berlin, Heidelberg, 1998. 73-85.

[5] Weiss, Karl, Taghi M. Khoshgoftaar, and DingDing Wang. "A survey of transfer learning." *Journal of Big data* 3.1 (2016): 1-40.

[6] Rawat, Waseem, and Zenghui Wang. "Deep convolutional neural networks for image classification: A comprehensive review." *Neural computation* 29.9 (2017): 2352-2449.

[7] Navneet Dalal and Bill Triggs - "Histograms of oriented gradients for human detection" Computer Vision and Pattern Recognition" 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 1. IEEE. 2005, pp. 886–893

[8] Gao, Mingliang, et al. "RGB-D-based object recognition using multimodal convolutional neural networks: A survey." *IEEE access* 7 (2019): 43110-43136.