A review on Face Recognition using Deep Learning Algorithm

Kevadkar Pragati Ashok[1], Gunjal Bhagyashri[2], Dhorje Rohini S[3]

[1]*Lecturer, Electronics & Telecommunication Department, Bhivrabai Sawant Politechnic, Wagholi*
[2]*Lecturer, Electronics & Telecommunication Department, Bhivrabai Sawant Politechnic, Wagholi*
[3]*HOD, Electronics & Telecommunication Department, Bhivrabai Sawant Politechnic, Wagholi*

---***---

Abstract— The face recognition system identifies a face by matching it with the facial database. Recognizing a face acquired from captured images or sensor images or sometimes taken from database images, or say real data for that matter is a complex task in itself due to the vast variations present in facial appearances and also because of the complexness of image background. In the present situation, face recognition is one of the amazing, efficient and widely deployed applications of image processing. Face recognition is used in many fields such as security, healthcare sector, as well as in, weather forecasting, pandemic detection marketing, disease prediction, self-driving cars, etc. Many methods are being developed to generate better results and accuracy in face recognition. But a deep learning approach has become a trend due to exceptional results and fast accuracy. This review paper gives an insight into the overview of some widely used deep learning schemes used under computer vision.

Keywords—Face Recognition, Artificial Intelligence, PCA, CNN, Machine Learning, Deep Learning.

1. INTRODUCTION

Detecting a face is a fundamental step and foremost problem in recognizing any pattern. It stands as a challenging and attractive area of computer vision. The utmost challenge in face recognition is to arise with competent feature representation to improvise the accuracy in different scenarios and provide improved results. Face recognition analyses the person's facial image input taken through a digital camera or face recognition done online. To generalize we can say that, face detection can be taken as a special type of object detection task under computer vision. For any given arbitrary image, the face detection has to tell whether there exists any face in the image, and if they are present, then send back the location of that image and extent of each face. Face recognition is that biometric method which has advantages of both low intrusive and highly accurate. Many algorithms are being put forward for face recognition. In recent times, deep learning has been able to achieve stupendous success in computer vision researches with significantly improving the state of art in classifying and recognizing the problems.

Face recognition algorithms developed on deep learning basics have achieved great when considered for processing time and accuracy. Deep Learning is a machine learning that adapts neural network architecture and consists of multi-layer perceptron from multi-hidden layers. By using a model architecture consisting of several nonlinear transformations, deep learning can find high-level features in the data. This feature is derived from the lower level to establish a representation depicting hierarchy. Deep learning is an enriched fraternity of methods that involves neural networks, hierarchical probabilistic models, and a good number of unsupervised and supervised feature learning algorithms.

The main advantage of deep learning practices is that they can be taught and trained using humungous datasets to learn the best features of representing data. In a simple task of face recognition algorithm, a face image is taken in the neural network to learn the face feature by making use of Convolutional Neural Networks (CNN), pooling layer and fully connected layer. These have to lead to the demand for computational resources at each step. Face recognition methods based on CNN trained with datasets can achieve lesser rates of error as they are capable to learn features that are robust and well defined to the real-time variations existing in the images of the face used at the time of training.

2. OVERVIEW

A. *Face recognition*

Face recognition has become an actively engaged research area across various disciplines such as pattern recognition, image processing, neural network, neuroscience. It happens to be a dedicated process and not just an application of the general object recognition process. Face recognition stands out as a biometric technology that tries to form an individual's identity. Face detection operates using a computer application that uses a digitally captured image of an individual's face (which at times can be taken from a video frame) and then examines it by comparing it to images of a database of previously stored records.

B. Face Recognition Structure

A generic face recognition system has the following three major steps:-

1. Gaining of face data,
2. Extracting face feature and
3. Recognition of a face.

Figure 1 shows the structure of where the object under consideration which will be provided to the system for recognition purposes is the acquired facial image.

Fig. 1. Block Diagram of the face recognition system

After that feature is extracted from the image and finally it is given for the recognition purpose. These steps are explained as follow.

(i) Gaining of Face Data

The first step is the acquiring and processing of face data. In this step, facial images are acquired from various sources. The source can be a camera or any easily available face image from the website database. The acquired face images must possess the pose, illumination, and expression which stands to be the variation so that the performance of the system can be analyzed under these conditions. The processing of the face database requires some time. Therefore input image is normalized and some image transformation methods are applied to the input image.

(ii) Extracting Face Feature

It can be defined as the process of taking out relevant and only useful information from a face image. In extracting a feature, a biometric reference which is a mathematical equivalent of its original image is generated, which is usually kept in the database and now this forms the basis or says vector of any task of recognition. Later these extracted features are used for recognition. The initial feature is a grayscale pixel.

(iii) Recognition of Face

Classifying the image becomes the next step when the features are extracted and selected. Various types of classification methods are used based on appearance in face recognition algorithms.

3. FACE RECOGNITION METHODS

A. PCA with Artificial Neural Networks

PCA with the ANN method is proposed in [11] which acknowledges face image's features for which PCA is used for extraction. PCA stands for reduction method of dimensionality and which retains the majority of variations existing in the set of data. It uses the captured variations in the dataset for encoding the face images. Feature vectors are computed. It then calculates the average of the face and it will hence normalize each input face image by reducing it from the average or mean face then computing the covariance matrix for it, and thus calculates covariance matrix's eigenvalues and keeps only the largest eigenvalues, then computes the eigenvector for the covariance matrix. By making use of that matrix eigenface is computed contacting the highest information of the facial image according to that it will compute the projected image.

PCA method reckons the maximum variations in data by following a conversion from a higher dimensional to lower-dimensional image space. Now the Artificial Neural Networks further processes these extracted projections of face images for training and testing purposes. For extracting the variations among the features of face images which contain the highest information with decomposed dimensions, PCA is used. Neural networks are trained using eigenfaces computed from extracted features and giving it as input to the Artificial Neural Networks. For the purpose testing, the trained neural network receives input from the eigenface of the tested image and it accordingly finds the desired match considering the threshold value for not accepting the unknown face and non-human images. Back Propagation feed-forward Artificial Neural Network (ANN) is utilized for training the input images of the face. The computed eigenfaces are fed to the neural networks. The number of different input face images decides the number of neural networks is taken. As we analyze that author has taken the 9 networks for 9 different face images. After setting the parameters neural networks are trained with eigenfaces of the input images via input layer, hidden layer, and output layer. Every eigenface image distance is examined and compared with every other image. The eigenfaces images of the same person have the zero distance between them and output is taken as 1 otherwise output taken as 0. The mathematical function values for each eigenface image are used to compare the eigenface images. In this work, the mathematical function Log-sigmoid is used for the eigenfaces of the same person, the specific neural network provides the output as 1 and for the eigenfaces of another person it provides the output as 0.Now, only the known faces are recognized as output 1. Neural Network forms an Identity matrix for different face images using the outputs 1's and 0's.

The errors in the output layer are given back to the previous layers and update the weights of these layers which minimize the error. The momentum and learning rate parameters count the updates from previous iterations and recalculate the newly updated output. For face recognition, feature extraction based on PCA is used for calculating the eigenfaces images of the test face image. This eigenface image is compared to each trained neural network. The Log-sigmoid function values are used to find the best match of tested eigenface by comparing it with the eigenfaces of the trained neural. As the threshold value is set which is the best distance. If the minimum distance between the tested eigenface image and the trained input eigenface image is less than the threshold value, then the output of the specific network is 1 and the trained eigenface image is selected from the Identity matrix as an output image and further recognized as a resulted face image otherwise the face image under test is rejected as non-human or unknown face image. The purposed face recognition system works with high accuracy and provides better success rates even for those face images which are noisy [11].

B. Linear Discriminant Analysis (LDA)

The linear discriminant analysis is a powerful method for face recognition. It provides an excellent representation that transforms linearly the subjected data into a low-dimensional feature space wherein the data is very well separated.

C. Convolutional neural networks (CNNs)

Convolutional neural networks (CNNs) are one of the most common types of deep learning procedures used for face recognition. The prominent advantage of deep learning methods is that they can be prepared by using huge amounts of data to learn a face representation that can deal with variations existing in the training data[1]. In this way, instead of designing customized features that can adapt to different types of intra-class variations (e.g. illumination, pose, facial expression, age), CNNs can understand them from data used for training purposes. The bigger setback with deep learning methods is that there exists a need for them to be trained with very huge datasets that contain enough differentiation to generalize to unseen samples. Fortunately, several big-scale face datasets have recently been released into the public domain that can be useful in training CNN models. Apart from learning discriminative features, neural networks can also work on reducing dimensionality and can be trained for using metric learning approaches or as classifiers. They are considered as completely trainable systems and those that have no requirement to be combined with any other methods.

CNN models for FR can be trained with the help of various approaches. One of them is considering the problem as a problem of classification, where each subject in the set of training is corresponding to a class. Once the training is done, the model can be made of use to recognize subjects that are absent in the set of training by removing the classification layer and using the features of the former layer as the face representation. Deep learning literature, refers to these features as bottleneck features. Following this first stage of training, the model can now be trained by deploying other techniques to optimize the bottlenecks for the exact application. Another usual approach in learning face representation is to learn directly the bottleneck features.

A complete face recognition CNN was proposed in[6]. This method makes use of a contrastive loss function. It implements a learning method that aims to reduce the distance between pairs of feature vectors that correspond to a similar subject while maximizing the distance between pairs of feature vectors that correspond to varied subjects. The CNN architecture used in this method was shallow and was trained with significantly small datasets. Facebook's DeepFace [4], one of the first CNN-based approaches for face recognition that used a high capacity model, achieved an accuracy of 97.35. Two novel contributions were made in this work: (i) an effective face alignment system based on 3D modeling of faces, and (ii) a CNN architecture containing locally connected layers [7] that (unlike regular convolutional layers) can learn different features from each region in an image. The three main factors that resultant in affecting the accuracy of CNN-based methods for face recognition are: training data, CNN architecture, and loss function. As in most of the deep learning applications, big training sets are required to prevent overfitting. Generally, CNNs which are trained for classification tend to become more precise as the amount of samples per class goes on increasing. When being exposed to more intra-class variations, the CNN model can learn more robust features. However, in face recognition, we are inclined towards extracting features that generalize to subjects which are absent in the data used at the time of training. Hence, the datasets need to have a large number of subjects due to which the model has exposure to more inter-class variations.

CNN consists of three main types of neural layers, which are, (i) convolutional, (ii) pooling, and (iii) fully connected layers. Each layer has a different role to play. Every layer transmutes the input volume to an output volume of neuron activation, which eventually leads to the fully connected layers, which provides results as the input data is mapped into a 1D vector. CNN's have been hugely successful in applications related to computer vision, such as self-driving cars, empowering vision in robotics, object detection, and face recognition.

- **Convolutional Layers**

In these layers, CNN uses various kernels to convolve the entire image as well as the intermediate feature maps, which generates various feature maps. Because of the benefits of the convolution operation, it is proposed as an alternative for completely connected layers to achieve faster learning times.

- **Pooling Layers**

These layers are involved in the reduction of the spatial dimensions of the input for the next convolutional layer. The pooling layer does not affect the depth of the volume. The operations performed are also called downsampling, as the reduction in the size leads to a simultaneous loss of data. But, such a loss stands to be profitable for the network since the decreased size makes way for reduced computational overhead for the following layers of in-network, and also this works against overfitting. Max and Average pooling are the most usually deployed strategies.

- **Fully Connected Layers**

Following numerous previously told layers, the high-level reasoning in the neural network is conducted via full connected layers. Neurons in this layer have complete connections to every activation in the former layer, as their name suggests. Their activation can thus be calculated with the help of a matrix multiplication which is then followed by an offset bias. Fully connected layers convert the 2D feature maps into a 1D feature vector. The derived vector could be either fed forward into a certain number of categories for getting classified or it can be assumed as a feature vector for conducting the processing further.

The architecture of CNNs employs the following three significant ideas: (i) local receptive fields, (ii) tied weights, and (iii) spatial subsampling. Based on the local receptive field, each convolutional layer unit is entitled to receive inputs from a set of the adjacent unit that belongs to the former layer. Thus now neurons can extract edges or corners which are elementary visual features. Then these features are combined by using subsequent convolutional layers that detect features of higher-order. In simple terms, in a convolutional layer, the units are organized in planes. So, all the units in one particular plane tend to share equal weights. Thus a specific feature is constructed by each plane. And these plane outputs are called feature maps. Each convolutional layer involves various planes, which helps in constructing numerous feature maps at each location.

One of the problems that may show up in the training of CNNs is that it has to deal with the huge amounts of parameters that have need to be learned, which may lead to overfitting which causes the problem. Techniques such as data augmentation, stochastic pooling, and dropout have been proposed. Furthermore, CNN's are usually subjected to pre-training, which is a process that initializes the network with pre-trained parameters and not the randomly set parameters. These can enhance the entire learning process and also accelerate the generalization capacity of the network.

As one method, the Convolution Neural Network (CNN) also has several pros and cons. Here are the advantages and disadvantages of the Convolution Neural Network (CNN)[8]:

Advantages:

1) It can be utilized in various image resolutions.
2) Computing is so detailed that the rate of error is likely small.
3) Convolution Neural Network (CNN) can solve problems that have a high complexity that has many parameters to be computed.
4) It can classify the face shape of known and unknown data.

Disadvantages:

5) Not suitable to be simple
6) Process long enough
7) Computing is very complex, directly proportional to the difficulty of the problems encountered.
8) Can't describe on the face with a certain position.

The accuracy level of the Convolution Neural Network (CNN) is quite high.

D. Convolutional neural networks (CNNs) with Pretrained models

In the convolution neural network, representation of face affects extensively the performance of the FR system and has also become a center of attention in the current research. We try to understand the study, which has employed a pre-trained convolution neural networks. This network is AlexNet. This pre-trained CNN network has been used to extract required image features and use them in the classification stage.

- AlexNet

AlexNet, brought by Krizhevsky et al. [13], was the first CNN to win the ImageNet challenge held in 2012, with a top 5 error of 16.4%. AlexNet introduced the use of rectified linear units (ReLUs). It includes 5 convolutional, 3 max pool, and 3 fully connected layers. This architecture makes use of [227 × 227 × 3] resolution image to be fed as input. In AlexNet, a 4096-dimensional feature vector represents the 227×227 image.

There are two approaches followed:

9) Application of the pre-trained CNN for feature extraction and support vector machine (SVM) for classifying.
10) Application of transfer learning from the AlexNet model for and extracting features.

The following stages are followed in the study. First, is the pre-processing stage, in which each image is resized to a considerable size for every CNN model and converted from a gray image to an RGB image. In the following stage, two pre-trained convolution neural networks are employed.

CNN networks have been used to extract relevant images of features and utilize them in the following classification stage. Finally, the process of classification of faces occurs with different convolutional neural networks. First, we used a pre-trained convolution neural networks, AlexNet for feature extraction, followed by SVM classifier. Next, transfer learning is applied from the previously trained AlexNet CNN for the classification task. Different datasets were used to conduct tests. Different results are looked at and the effectiveness of each approach is analyzed and the results are compared when transfer learning from pre-trained AlexNet and using support vector machines (SVM). SVM as a classifier is used to identify faces because it has an observable classification result on data which is not linear. SVM has tremendous advantages in solving problems as pattern recognition and machine learning for FR function overfitting.

The SVM [14] classification is a process where a supervised binary classification method is used and when a training set is introduced, wherein the algorithm develops a hyperplane that maximizes the margin that exists between two input classes. For instance, considering linearly separate data with two distinct classes, the system can have numerous hyperplanes that separate two classes. What SVM does is that it identifies the most ideal hyperplanes that have a maximum margin between all available hyperplanes, whereby the margin is the distance difference between the hyperplane and the support vectors. However, in the real world, the data are not always linear, and it is not possible to classify by a linear classifier, and thus the non-linear SVM classifier is proposed.

- VGG-Face Network

A VGG-Face or deep convolutional network [15] is proposed for face recognition by using the VGG Net structural design [16]. It is trained using approximately 2.6 million images of the face of 2522 individuals derived from the internet. Its network involves 16 convolutional network layers, 3 fully connected layers, 5 max-pooling layers, final linear layer which has Softmax activation. It takes the color image as input covering the size of 224×224 pixels and makes use of regularization in the fully-connected layers. On top of that, it relates ReLU activation to wholly convolutional layers. VGG network which crosses 145 million parameters proves that it is surely a computationally expensive design.

4. CONCLUSION

Face recognition study stands to be a challenging domain for researchers for many past years. This paper tries to review a significant amount of papers to cover up the past advancements in the area of face recognition. The present study reveals that the methods and technology for face recognition have grown and developed over time from PCA, LDA to CNN. And in today's time, CNN happens to provide better results even for the images which have variations.

5. REFERENCES

[1] Daniel S´aez Trigueros, "Face Recognition: From Traditional to Deep Learning Methods", Li Meng School of Engineering and Technology University of Hertfordshire Hatfield AL10 9AB, UK
[2] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," arXiv preprint arXiv:1411.7923, 2014.
[3] O. M. Parkhi, A. Vedaldi, A. Zisserman, et al., "Deep face recognition.," in BMVC, vol. 1, p. 6, 2015.
[4] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1701–1708, 2014.
[5] H. Fan, Z. Cao, Y. Jiang, Q. Yin, and C. Doudou, "Learning deep face representation," arXiv preprint arXiv:1403.2802, 2014.
[6] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 1, pp. 539–546, IEEE, 2005.
[7] G. B. Huang, H. Lee, and E. Learned-Miller, "Learning hierarchical representations for face verification with convolutional deep belief networks," in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pp. 2518–2525, IEEE, 2012.
[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Proceedings of the 26th Annual Conference on Neural Information Processing Systems (NIPS '12), pp. 1097–1105, Lake Tahoe, Nev, USA, December 2012.
[9] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proceedings of the 27th IEEE Conference on ComputerVision and Pattern Recognition,pp.580– 587, Columbus, Ohio, USA, June 2014.
[10] M. M. Kasar, D. Bhattacharyya, and T. Kim, "Face Recognition Using Neural Network : A Review," vol. 10, no. 3, pp. 81–100, 2016.
[11] N Jindal, V Kumar," Enhanced Face Recognition Algorithm using PCA with Artificial Neural Networks", International Journal of Advanced Research in Computer Science and Software Engineering, vol 3, no. 6, (2013), pp. 864-872.
[12] H Sahoolizadeh, D Sarikhanimoghadam, and HDehghani," Face recognition system using a neural network with Gabor and discrete wavelet transform parameterization", International Conference of Soft Computing and Pattern Recognition, Tunis, (2014), pp. 17 – 24.
[13] Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, CA, USA, 3–6 December 2012; pp. 1097–1105.
[14] Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A training algorithm for optimal margin classifiers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, PA, USA, 27–29 July 1992; pp. 144–152.
[15] Yosinski J, Clune J, Bengio Y, Lipson H (2014) How transferable are features in deep neural networks? In: Advances in neural information processing systems, pp 3320–3328
[16] Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556