Face Recognition and Identification Using Deep Learning

Hareendran Ullattil

Lecturer in Electronics Engineering.
Government Polytechnic College, Meenangadi, Wayanad, Kerala.

***Abstract*: Face recognition is one of the most challenging field of image analysis and computer vision due to its wide practical applications in the areas of biometrics, information security, law enforcement and surveillance systems. It has been a topic of active research proposing solutions to several practical problems giving rise to the significant amount of research in recent times aimed at addressing the challenges of face recognition attributed to the following factors such as illumination, emotion, occlusion, facial expressions and poses, which greatly affect the performance in achieving efficient and robust face recognition systems. In this field, many researchers adopted different techniques that solely rely on extracting handcrafted features to achieve better results. Recent development in deep learning and neural networks have made it possible to achieve promising results in numerous fields including pattern recognition and image processing. Deep learning methods boost up the learning process and facilitates the data creation task. Many algorithms have been developed to use deep learning architectures to get maximum result and achieve the state-of-the art accuracy. Some algorithms design their architectures from scratch and others fine-tuned the existing models to get maximum efficiency of generalization power. Algorithm complexity, data augmentation and loss minimization are the main concern of deep learning paradigms. We have reviewed these architectures in relation to algorithm complexity and experimental results on benchmark dataset.**

***Keywords*: Face Recognition, Deep Learning, Face Identification, Face Verification**

Introduction:

Face recognition has been one of the most actively studied topic in the computer vision community. With the advancement in technology and the high usage of multimedia application in smartphone, the challenge for face detection and efficient recognition is greater than before. There has been great advancement in face recognition, starting with the Viola Jones as pioneer work for detecting frontal-face in real time along with low computational complexity. This was followed by different approaches that involved basic image processing techniques that would extract various facial features from the face images and were fed to different classifiers for training and recognition. Other than this most of the initial approaches in the area of face recognition used up-right images without much variations in pose, illumination, occlusions
etc. Although the initial approaches worked well for the front-face images but failed with different angles or illuminations. The various other classifiers used lack the ability to classify multi view facial features. This led to other approaches for multi-view face recognition and approaches that would eliminate the hand crafted facial features. [1]

Face recognition can be broken down into two steps stated below:

Step 1: Identification

Recognizing individual by locating their faces in a given image is the first step in Face recognition system. The identification ensure that the algorithm identifies the image as a facial image and then utilizes this information to identify the faces in the image. The identification step checks for the face in the image against the other faces to look for the identity of the face in the image, which makes this a multiclass classification problem.

Step 2: Verification

The verification is concerned with validation of identity based on the input image of a face. It performs a one-to-one matching by either accepting or rejecting the identity which makes this a binary classification problem.

Deep learning:

machine learning methods that is inspired by the neural networks. It involves different models based on the neural networks and it uses multiple layers for feature extractions. Its layer sequence or architecture is such that each layer input is an output from the previous layer. It learns in a supervised as well as unsupervised manner. Deep learning has many models like Deep Neural network, recurrent neural network and convolutional neural networks. These models have improved the ability of classification, recognition, detection and localization. Deep learning is now being advancing by the development of new machine learning approaches, new versions of neural networks and increasing computational powers like GPU. [2]

Deep learning is a product of development of artificial neural network. At the beginning, practice of training MLPs (Multi-layers Perceptron) where a linear layer is added from input of network connection to that of output. Subsequently, G.

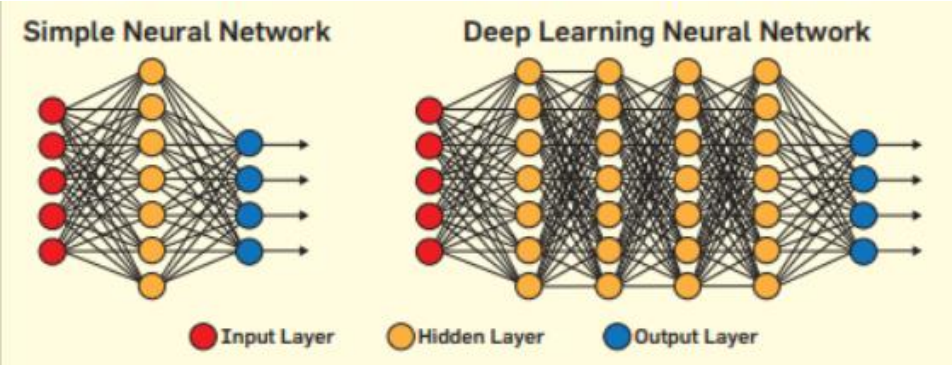

Figure 1: Simple neural network vs deep learning neural network

Thomson had proposed a new idea known as deep learning, where it is a new model training as shown in Figure 1.

Applications of deep learning:

1) Image recognition: Deep learning has been widely used in computer vision and it has shown promising results in this area. Practical problems and many computer vision challenges are now being addressed with deep learning producing better results than humans.

2) Speech Recognition: Deep learning has shown convincing results on speech recognition. For this purpose, recurrent neural networks are used. Today all commercial speech recognition system like Baidu, Google now and Skype translator are based on deep learning.

3) Natural Language Processing: LSTMs a recurrent network has been using for many natural processing tasks like sentiment analysis, word translation and many more and they are producing outstanding results on these problems.

4) Recommendation Systems: Recommendation systems are using deep learning to learn user interest and preferences and to recommend their interests.

Flowchart of the face recognition:

To do face recognition, there must be an input to be detected and verified. Hence, an image sensor or typically a camera has to be set up for recording or capturing images. The camera should be compatible with the software used. The next step is the input image. The input can be images and recorded video or real-time video. After the input is provided, faces in the images or videos are to be detected. When the classifier is trained, it can be utilized to start to recognition work. It can be used in either video or image to recognize one or more person. Different set of python scripts are provided to run the different type of recognition. The python script will import the classifier that is trained in previous step in order to carry out the recognition for the person from the camera or from an image. [3]

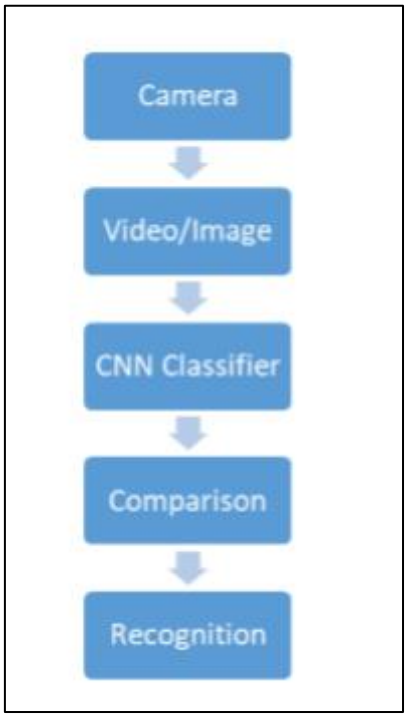

Figure 2: Flow chart of face recognition

Figure 2 shows the flowchart of the face recognition steps.

In face detection, Haar feature-based cascade classifiers is used and the classifier used is Haar Cascade for frontal face. A Haar Cascade is basically a classifier which is used to detect the object for which it has been trained for, from the source. The Haar Cascade is performed by superimposing the positive image over a set of negative images. The training is generally done on a server

and on various stages. Better results are obtained by using high quality images and increasing the amount of stages for which the classifier is trained.
TensorFlow is the framework that is being used in the system classifier section. Classifier is trained and used in the recognition process. The training process takes a long time to achieve a better classifier. The longer the time of the training runs, the better the classifier is. In the proposed face recognition system, the training period taken is 3 days. If the training is allowed to run longer, the loss can be reduced further and hence the accuracy can be increased.

Objectives:

1. To find a series of data of the same face in a set of training images in a database.
2. To understanding of how people process and recognize each other's face, and the development of corresponding computational models.
3. To study of development of corresponding computational models.

Review Of Literature:

This paper focuses on the detection of Multi-view faces. De- spite many extensive studies, many techniques have been used but still require annotations of the facial landmarks in addition to multiple trained models to learn the faces in different orientations. This paper proposes a Deep Dense Face Detector that does not need to annotate the faces and it can detect faces in many different orientations by just using a single model based on deep convolutional neural network.
A. Krizhevsky et al proposed model that uses a pre-trained. It is a CNN model that took part in ILSVRC- ImageNet large scale visual recognition challenge; they just fine tuned it for their face detection problem. They trained the model on Annotated Facial Landmarks in the Wildm(AFLW) dataset that contains 21K images with 24K face annotations.[4]
S. Matuska et al. did the comparison of the speed between OpenCV and Matlab. Basic algorithm of image processing is presented and the time consumption in OpenCV and Matlab. It is experimented that OpenCV is much faster than Matlab up to 30 times and can be up to 100 times for Erosion algorithm. [5]
Caricatures [Brennan 1985; Bruce 1988; Perkins 1975]: A caricature can be formally defined [Perkins 1975] as "a symbol that exaggerates measurements relative to any measure which varies from one person to another." Thus the length of a nose is a measure that varies from person to person, and could be useful as a symbol in caricaturing someone, but not the number of ears. A standard caricature algorithm [Brennan 1985] can be applied to different qualities of image data (line drawings and photographs). Caricatures of line drawings do not contain as much information as photographs, but they manage to capture the important characteristics of a face; experiments based on non ordinary faces comparing the usefulness of line-drawing caricatures and unexaggerated line drawings decidedly favor the former [Bruce 1988]. [6-8]
Distinctiveness [Bruce et al. 1994]: Studies show that distinctive faces are better retained in memory and are recognized better and faster than typical faces. However, if a decision has to be made as to whether an object is a face or not, it takes longer to recognize an atypical face than a typical face. This may be explained by different mechanisms being used for detection and for identification. [9]
Movement and face recognition [O'Tooleet al. 2002; Bruce et al. 1998; Knight and Johnston 1997]: A recent study [Knight and Johnston 1997] showed that famous faces are easier to recognize when shown in moving sequences than in still photographs. This observation has been extended to show that movement helps in the recognition of familiar faces shown under a range of different types of degradations—negated, inverted, or thresholded [Bruce et al. 1998]. Even more interesting is the observation that there seems to be a benefit due to movement even if the information content is equated in the moving and static comparison conditions. However, experiments with unfamiliar faces suggest no additional benefit from viewing animated rather than static sequences. [10-11]

Research Methodology:

In this paper, we presented a literature survey of latest advances in researches on machine learning for face recognition and their experimental results on public databases.
Books, educational and development journals, government papers, and print and online reference resources were only some of the secondary sources we used to learn about the composition, use, and impacts of Face Recognition and identification using deep learning.

Result and Discussion:

1.Distance of Face Detection:

When the distance is close or less than 60cm, the proposed system can barely detect the face. In the other hand, when the distance is extended to more than 60cm, the recognition takes place.

Figure 3: Distance between face and camera <60cm

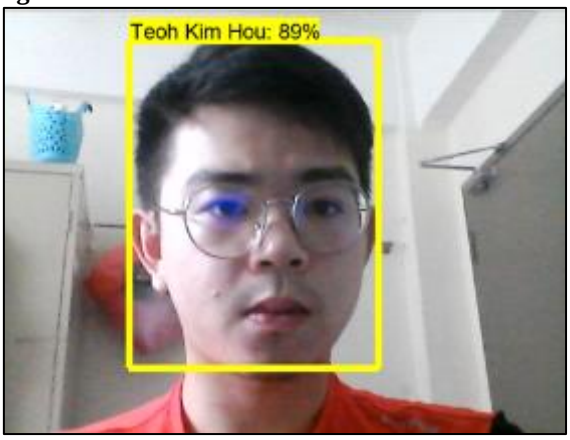

Figure 4 : Distance between face and camera >60cm.

Figure 3 and 4 show the distance between face and camera is affecting the recognition process. Accuracy of Face Recognition based on Image Multiple photos either in group or individual are loaded into system to verify the accuracy. A person should had appeared in those photos for 20 times. When the photos are all tested with the proposed face recognition system, the data is computed in confusion matrix to calculate the accuracy of the system. [12]

2. Accuracy of Face Recognition based on Image:

Multiple photos either in group or individual are loaded into system to verify the accuracy. A person should had appeared in those photos for 20 times. When the photos are all tested with the proposed face recognition system, the data is computed in confusion matrix to calculate the accuracy of the system.

Table 1: Confusion matrix for image recognition

Number of Face Recognition			Result
Person 1	Person 2	Person 3	
17	1	2	Person 1
1	18	1	Person 2
0	0	20	Person 3

From Table 1, it can be observed that the true and false recognition done by the proposed face recognition system. For the first person, 17 out of 20 recognitions are true. The true statement means that the identity of person in the photo that is recognized by the system is matched with real identity of the person. For second person, 18 of his photos are recognized correctly while for the last person, all photos are recognized correctly. Thus, the accuracy of the system can be calculated. [13]

3. 3D Model of the Face Image:

Face recognition usually consists of four stages, detect face align represent classify. revisited the align and represent stages by doing 3D face modelling and applying transformations and rotations to align the face better, and then it represents the face using nine-layer deep neural network. [14]

Given below are the stages for face recognition employed.

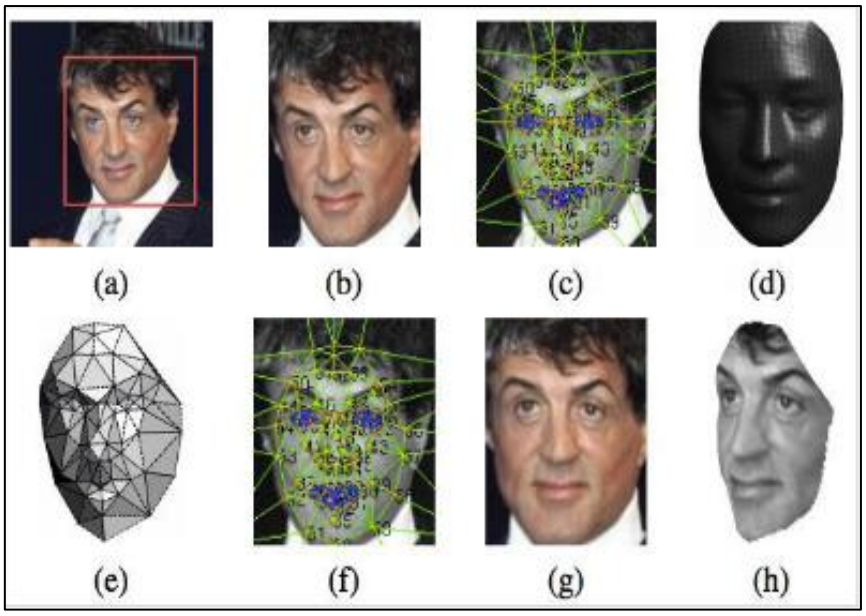

Figure 5: 3D Model of the Face Image

In order to align the face, paper has used 3D model of the face. Initially it has detected six reference points in the image and on the basis of these reference points it has cropped the face, then it has marked 67 reference points and on the basis of it 3D mesh has been created. From 3D mesh 2D image has been created using affine transformations. Result (g) would be the input to the network. 3D-algined image (152 by 152 pixels) is an input to the network which is passed through a convolutional layer, a max pooling layer, and a convolutional layer. They did not add many pooling layers because they believe that pooling layers remove information about the face. Next three layers in the network are locally connected layers. Network has two fully connected layers.

Conclusion:

This paper has reviewed the latest studies to provide a good knowledge of successful growth of deep learning in the field of face recognition. We have seen that different deep learning techniques has performed outstandingly on benchmark datasets like LFW and YouTube Faces (YTF), but the results of "FaceNet: A Unified Embedding for Face Recognition and Clustering" has outperformed other studies with its outstanding architecture. It has been seen that deep learning models perform better when they are trained with a large dataset. In future face recognition can further be improved using deep learning by tweaking the best studies and by using different data augmentation techniques that will generalize the face recognition model.

From the results, it can be concluded that the accuracy of face recognition on image is higher than the accuracy on the real-time video. It can be observed that the resolution of image is a lot better than the real-time video. Due to the limitation of processing system, the fps of the video is low and causing the face captured is not as clear as in the photo. Hence, the classifier tends to assume that a particular person exists the characteristic of the other person. Therefore, the accuracy of the real-time video recognition has been reduced to 86.7%.

Reference:

1. P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," arXiv preprint arXiv:1312.6229, 2013
2. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097-1105.
3. S. Zhaoqing, Z. Su, and L. I. Zhicheng, "Face Images Recognition Research Based on Smooth filter and Support Vector Machine *," pp. 2760–2764, 2010.
4. A.Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097-1105
5. S. Matuska, R. Hudec, and M. Benco, "The Comparison of CPU Time Consumption for Image Processing Algorithm in Matlab and OpenCV," pp. 75–78, 2012
6. BRENNAN, S. E. 1985. The caricature generator. Leonardo, 18, 170–178.
7. BRUCE , V. 1988. Recognizing faces, Lawrence Erlbaum Associates, London, U.K
8. PERKINS, D. 1975. A definition of caricature and recognition. Stud. Anthro. Vis. Commun. 2, 1–24
9. BRUCE , V., B URTON, M., AND D ENCH, N. 1994. What's distinctive about a distinctive face? Quart. J. Exp. Psych. 47A, 119–141
10. O'T OOLE, A. J., ROARK, D., AND A BDI, H. 2002. Recognitizing moving faces. A psychological and neural synthesis. Trends Cogn. Sci. 6, 261–266.
11. K NIGHT, B. AND J OHNSTON, A. 1997. The role of movement in face recognition. Vis. Cog. 4, 265–274.

12. N. K. Balcoh, M. H. Yousaf, W. Ahmad, and M. I. Baig, “Algorithm for Efficient Attendance Management: Face Recognition based approach,” vol. 9, no. 4, pp. 146–150, 2012.
13. S. Matuska, R. Hudec, and M. Benco, “The Comparison of CPU Time Consumption for Image Processing Algorithm in Matlab and OpenCV,” pp. 75–78, 2012
14. W.-S. T. WST, "Deeply learned face representations are sparse, selective, and robust," perception, vol. 31, pp. 411-438, 2008