International Journal of Advance Research in Science and Engineering
Vol. No.6, Issue No. 09 , September 2017
www.ijarse.com
IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

Comparative Evaluation of Deep Architectures for Face Recognition in Unconstrained Environment (FRUE)

Deeksha Gupta

Department of Computer Science and Applications,
MCM DAV College for Women, Chandigarh, (India)

ABSTRACT

Face recognition is one of the most reliable and popular recognition schemes under biometric systems. Face recognition methods perform well on the dataset where all face images are carefully collected and are frontal and free from noise. But for face images suffering from change in illumination, brightness, background, expression, pose etc., the face recognition becomes a challenging task. Deep learning has carved a niche in developing artificial intelligence based applications by providing high quality results. Three most important requirements for a reliable Deep learning based Face recognition system is large training dataset of face images, an effective training procedure and a powerful validation mechanism. In this paper current state of art deep architectures for face recognition are investigated and evaluated. This paper also specifies various training and testing datasets for face recognition with their availability.

Keywords: Face Recognition, Deep Learning, Convolutional Neural Network

I. INTRODUCTION

Face recognition is a technique in which computers are used for the analysis of the face to distinguish the biological features. It is one of the most fast growing areas of research in image processing and computer vision due to its numerous real time applications. Its wide range of applications include security system, access control, Surveillance, image database investigations, entertainment, social websites ,mobile applications and Pervasive Computing[1]. Though face recognition is the basic ability of human beings to identify and recognize human faces but it is a difficult problem for computers to identify and verify faces especially in unconstrained environment. The reason behind intricacy is large variation in the face expression, illumination, background, brightness, partial occlusion, alignment of camera, pose and facial change due to aging of the person. In spite of major recent advances in the field of face recognition it is always a challenge to meet human being recognition potential in case of unconstrained conditions.

Nowadays the researcher's focus in face recognition research has moved from traditional face recognition systems based on geometric characteristics or statistical characteristics like PCA and LDA based algorithms to deep learning[2]. Deep learning application has already shown its success in various fields like speech recognition, natural language processing, object detection and identification and many more. Deep learning facilitates a general purpose learning procedure to learn highly distinctive features automatically. The biggest command of deep learning is its capacity to classify the given data with high accuracy and deriving generic conventions. Deep learning makes the classification process sensitive to minute detail and robust to irrelevant

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

variations like illumination, orientation, brightness and background. Deep learning, also known as, Deep Convolutional Neural Networks (DCNN) are biologically inspired improvement over multilayer perceptron neural networks (MLPNN)[3].Although, deep networks are much harder to train as compared to simple neural network but outperform the later.

This paper includes extensive evaluation and comparison of various DCNN architecture devised for Face Recognition in Unconstrained Environments (FRUE). Section 2 of the paper specifies the various popular training and testing face images datasets, section 3 covers methodology behind face recognition using Deep Learning , section 4 represents various state of art Deep architectures for face recognition and section 5 contains the extensive evaluation Deep Learning based FR systems.

II. TRAINING AND TESTING DATA SETS

DCNN approach is a data driven approach. Larger training set makes training of DCNN a very time consuming affair but results in better face recognition performance. An effective training and testing dataset for FR systems should be generalized to unconstrained environment. The face images of dataset should be rich in variations in pose, expression, illumination, brightness, position of camera etc. Various popular training and testing dataset with their credentials are given in Table 1.1.

Training and Testing Dataset			
Dataset Name	**#images**	**#unique identities**	**Availability**
WDref [4]	99773	2995	Public
CASIA WebFace[5]	494,414	10,575	Public
CACD[6]	163,446	2,000	Public
FaceScrub[7]	1M	5,30	Public
CelebFaces+[8]	202,599	10,177	Private
SFC (Social face Classification)[9]	4.4M	4,030	Private
VGG-Face[2]	2.6M	2,622	Public
Google[10]	200M	8M	Private
IDL DB[11]	1.2 M	18,000	Private
LFW(Labeled Face in Wild [10][9][8]	13,233	5,749	Public
YTF[9]	3,425 videos	1,595	Public
MegaFace[12]	1M	690,572	Public
IIB-A	5,712 images 2,085 videos	500	Public

Table 1.1-Training and Testing Face Datasets

For face recognition, DCNN is trained against all the images in dataset. Feature vector is extracted for all the images of training dataset and stored in database. The various parameters used for training of DCNN are adjusted using back propagation and appropriate optimization technique.

III. METHODOLOGY

Face recognition system is a multistage process. Broadly classified, the various stages of FR system include image preprocessing, feature extraction, image classification followed by face verification shown in Fig. 1. In

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

Deep learning based Face recognition the feature extraction and image classification is performs using Deep Convolutional Neural Network whereas image processing may include deep learning based procedures.

Figure 1: multi-stage Face recognition process

3.1 Image Preprocessing

This is the first step of any face recognition system. In this step the image is made ready for image representation. This stage has three basic components: Face detection followed by Image cropping and Face alignment.

3.1.1 Face detection

Face detection or face localization is the first most step of image preprocessing. Some Face Recognition systems don't need face detection because the images in training dataset are already in normalized form. However, Algorithms used for face recognition in unconstrained environment require to perform face detection. Face detection includes identification of sub-regions in image containing face, ignoring the background and other objects in the image and enclosing such sub-regions in bounding box. Face detection algorithms also perform face normalization for further effective face analysis of image[13]. Face normalization deals with elimination of irrelevant information for face identification. It includes the removal of noise, illumination, brightness from image as well as normalizing the face image size. The performance of face detection method is measured in terms of detection rate and false alarm. Face detection methods can be either sub space based or feature base methods[10] [13]. But in case of face detection in unconstrained environment sub-space based and features based methods fail to confine the discriminative information whereas CNN based detection algorithms are providing good results. So overcoming from this limitation of traditional face detection methods, in past decade CNN based face detection models [14] [15] [16] are getting more popularity.

3.1.2 Face Alignment

After face detection face alignment is performed, where the required 2D or 3D transformations are performed on face image depending upon the requirement of deep model architecture. The vital requirement of face alignment process is to localize the characteristic fiducial Points (two eyes, nose and mouth corners) in the cropped face image, so that transformation can be done corresponding to the localized positions.

Schroff et. al [10] proposed FR system where no or little 2D transformation is required. FR methods given by *Wang* [8] and *Liu* [11] perform 2D affine transformation for face image alignment. *Taigman et. al* [9] proposed method performs both 2D and 3D affine transformations to generate 3D frontal view model of face for further processing.

3.2 Image Representation (Feature Extraction)

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

It is the second stage of face recognition system which includes extraction of sophisticated features from preprocessed image. In deep learning the bottom layers are used to extract low level features from the image like presence of edges, corner, texture and moving towards top layers the high level features like position and size of face, nose, eyes mouth etc are mined from low level output features of previous layers. The subsequent convolutional layer combines the features of previous layer using formula given below:

$$(y)^{j(r)} = max\ (0\ , b^{j(r)} + \sum\ k^{ij(r)} ** x^{i(r)}$$

$x^i = i^{th}$ input map, $y^j = j^{th}$ output map, k^{ij}=convolutional kernel between map i and j, b^j=bias of j^{th} output map.

Various filters like sobel operator[17], Gaussian filter[9], Gabor filters[10], Haar filter[16] are used to extract features from image. These extracted features are represented in the form of feature vector. As architecture moves from bottom layers to top layers the dimension of feature vectors is reduced intentionally to reduce the computational cost.

3.3 Classification

This stage is responsible to label the image with the corresponding class with some tolerance error rate. At this stage the features extracted at image representation stage are used to classify the image. Here the high dimensional image feature vector is down sampled to low dimensional sophisticated feature vector and using fully connected layer with classifier the aim is achieved. One to many image mapping is performed to find the corresponding class for the input image. The various classifiers available are SVM[9][10], Softmax[5][9] [8]etc.

3.4 Face Verification

Face verification is second paradigm of face recognition system. Face verification ensure that two input images of same identity should fall under same class. Under face verification one to one image mapping is done. Metric learning method is used to verify the face with the face image database. Weighted Chi square[9], Joint Bayesian[8][14],Triplet loss based on L2 distance[10], cosine function[5] are some example of metric learning methods that can be used for verification. Verification methods are independent on feature extraction procedures and can process both hand crafted and CNN learned features.

IV. DEEP ARCHITECTURES FOR FACE RECOGNITION

4.1 Deepface

Taigman et. al [9] proposed method where the 3D aligned face images are generated to cope up with out of plane rotations, using multi stage approach, to train the DCNN against large dataset SFC. In [9], First, the input image is cropped to face followed by 2D transformations. After performing 2D alignment, 3D affine transformations are done to produce 3D aligned front face view using 3D analytical model. The resulting 3D align images are given as input to DNN. The DNN has 8 layers architecture. In this architecture after 3 layers of interleaved convolutional layer and pooling layer, three locally connected layers are used followed by 3 fully connected layers. The purpose of locally connected layers is to generate efficiently the high level features without adding to the computations. In [9], The most of the generated features are sparse due to application of nonlinear Function ReLU after every convolutional layer, Local connected layer and fully connected layer and Dropout Regularization method. To train the DCNN, cross entropy loss is minimized by computing the gradient

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

using back propagation and updating the parameters by SGD. K-way softmax classifier (k is no. of classes) is used for classification after the last layer of architecture. The author tested the proposed system against LFW and obtained remarkable performance (97.35%).

4.2 DeepId

Wang [8], proposed CNN based face recognition architecture, named as Deep hidden IDentity (DeepID), consists of network fusion of CNNs. In Wang [8] face recognition approach, During preprocessing stage the face images are aligned across 5 fiducial points(center of 2 eyes, mouth corners, nose tip) using 2 D transformations. Then the input image is segmented into 60 patches to extract different features from different parts of face image. Each patch is of size 31X31X1 or 39X31X1. These patches are given as input to DCNN. The network architecture is made up of 9 layers. First seven layer are interleaved convolutional layer and pooling layers. The eighth layer is DeepID layer and last layer is Softmax layer. The length of output feature vector is reduced to 160 at fully connected deepID layer so that only very high level, compact and discriminative features can be used for classification and irrelevant variations can be suppressed. The DeepID feature is obtained by combining DeepId layer output feature of all networked CNN. For training of network SGD (Stochastic Gradient Descent) method is used with gradient calculated by back propagation method. For similarity metric learning Joint Bayesian method is implemented.

Successors of DeepId named DeepId2[18], DeepId2+ [19], DeepId3[20] were developed to improve the performance. Instead of using only identification information for DCNN training DeepID2 applies both identification and verification signals for training of DCNN[18]. DeepId2+ further improves performance by using much larger training set (combination of CelebFaces+ and WDRef datasets), increasing the no. of filters and including Fully connect layer at early with supervisory information [19]. The latest version DeepId3 has even more denser CNN network and outperforms its predecessors.

4.3 FaceNet[10]

This face recognition method is produced by Google research group. In this approach training images are thumbnail that are cropped to face area, so no 2D or 3D transformation is required. This dense architecture has 22 layers. Along with convolutional layer, pooling layer, Fully connect layer, DCCN also contains local Normalization layers to perform normalization over local input regions. Normalization layer works on the volume of the input map vector. 140 M parameters are required for training of DCNN. DCNN generates 128D embedding (Euclidean distance) for image representation by using triplet loss function. The triplet selection is the most crucial component of this Face recognition system. A triplet consists of group of three face images where two images are of same person and one image of different identity. The triplet selection is done by using data mining techniques. Euclidean distance of images is generated such that the same identity images have less distance than that of different identity image. Unlike [8], only one CNN is implemented and Support Vector Machine classification function is used for mapping of input image to its corresponding class.

4.4 VGG Net FR

Parkhi et. al. [2] proposed a face recognition system that utilized VGG net model for designing the architecture. During preprocessing only 2D transformations are applied on face images. In VGG based model multiple

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

continuous convolution layer jointly are used to extract complex features. Normalization and ReLu implementations speed up the training of CNN. The contiguous convolution layers are used with small size kernel to reduce no. of parameter and hence the computational cost. Inspired from Schroff et. al.[10] , VGG Net FR has very deep CNN architecture and uses triplet loss based metric learning method. CNN is trained by optimizing the multinomial logistic regression using SGD (Stochastic Gradient Descent). Softmax log function is implemented for classification and L2 based metric learning method is used for face verification.

4.5 Webface

Dong et. al. [5] proposed another CNN based face recognition architecture, referred to as WebFace.The work in [5] includes collection of face image database containing around 494,414 images of 10,575 unique identities, which is made publicaly available, and training CNN for the collected database. Webface uses a much deeper CNN architecture and multiple loss functions.It has 17 layers architecture , having 10 convolutional layer, 5 pooling layer and 3 fully connected layers. inspired from VGG model [2], *Dong et. al.* [5] designed architecture uses 3X3 sized kernels to cope up with complexity of network. Like Wang [8], Webface also combines softmax identification and Contrastive verification loss functions as objective function.

V. EVALUATION

Above investigated deep architectures are evaluated on the basis of various factors related to underlying deep architecture and its training. These factors include training dataset used, required image alignment before CNN training, number of layers in underlying deep architecture, different layers with sequence in the architectures, input image size, patch fusion , no. of parameters required for training, length of output feature produced at last layer for driving classification, classifier used, accuracy obtained on LFW and metric learning method adopted for verification .

	Deepface	**FaceNet**	**DeepID**	**CASIA/webface**	**VGG Net**
Training Dataset	SFC	Google	CelebFace+	CASIA WebFace	VGG Face
Alignment	2D and 3D affine Transformation	No alignment	2D Transformations	2D Transformations	2D
#Layers	8	22	9	17	21
Architec-ture Layers*	C1→P2→C3→L4→L5→L6→FC7→FC8	C1→P2→N3→C4→CP5→N6→P7→C8→CP9→p10→C11→CP12→C13→CP14→C15→P16→P17→CC18→C19→FC20→FC21→	C1→P2→C3→P4→C5→P6→C7→DeepID8→ FC9	C1→C2→P3→C4→C5→P6→C7→C8→P9→C10→C11→P12→C13→C14→P15→FC16→FC17	C1→C2→P3→C4→C5→P6→C7→C8→C9→P10→C11→C12→C13→P14→C15→C16→C17→P18→FC19→FC20→FC21

		FC22			
I/P Image Size	152x152x3	220x220x3	39×31×{3,1} 31×31×{3,1}	100×100×1	220x220x3
Patched I/P	No	No	Yes	No	No
#Para-meters	120M	140M	101M	5015K	133 M
O/P Feature Length	4096	128	19200 (160/Patch)	320	1000
Loss Function	Softmax	SVM	Softmax	Softmax	Softmax log
Accuracy	97.35%	98.87%	97.45%	97.73%	98.95%
Verifica-tion metric	Weighted $\chi 2$ distance, Siamese network	L2	Joint Bayesian	Cosine	L2

Table 2 : Comparison Matrix

*Layers : C-Convolutional Layer followed by ReLu, Cp-Cross Channel Pooling Layer (network in network layer), P- Pooling Layer, N-Normalization Layer, L-Local Connected Layer, FC- Fully Connected Layer, CC-Concatenation Layer

VI. CONCLUSION

Deep learning has become breakthrough in face recognition. Deep Convolutional Neural Networks have shown tremendous reduction in the face recognition error rate due to their capability of learning from a very large training data set like SFC, LFW(Labeled Faces in Wild benchmark) and YTF (YouTube Face Dataset)[10] especially for face recognition in unconstrained environments (FRUE).

In this paper an extensive survey of various Deep learning based architectures for face recognition is given. Along with the architecture details of various methods their comparison is done on various factors like no. of parameters, feature length generated and their performance.

Although DCNN are performing very well in face recognition, there are some aspects of DCNN that require further investigation, as during training and testing DCNN requires large memory to store parameters and high time for implementation. Although SGD algorithms are performing very well with GPU based implementation but still some more effective, fast and scalable algorithms are required to develop.

Deep learning methods need to handle very large no. of input- outputs nodes, parameters and hyper-parameter (no. of layers, learning rate, kernel size, stride, padding size etc.). All these parameters and hyper parameters are internally related. So it is quite difficult to adjust them together. So improved optimization techniques need to be design to make training of Deep Convolutional Neural Network easier.

REFERENCES

[1] L. Xiong, J. Karlekar, J. Zhao, J. Feng, S. Pranata, and S. Shen, "A Good Practice Towards Top Performance of Face Recognition: Transferred Deep Feature Fusion," vol. XX, no. Xx, pp. 1–10, 2017.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

[2] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," no. Section 3, 2015.

[3] H. Shimodaira, "Multi-Layer Neural Networks," pp. 2–4, 2015.

[4] D. Chen, X. Cao, L. Wang, F. Wen, and J. Sun, "Supplemental material for ' Bayesian Face Revisited : A Joint Formulation ,'" no. 2, pp. 1–5.

[5] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning Face Representation from Scratch."

[6] B. C. Chen, C. S. Chen, and W. H. Hsu, "Cross-age reference coding for age-invariant face recognition and retrieval," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8694 LNCS, no. PART 6, pp. 768–783, 2014.

[7] H. W. Ng and S. Winkler, "A data-driven approach to cleaning large face datasets," *2014 IEEE Int. Conf. Image Process. ICIP 2014*, pp. 343–347, 2014.

[8] X. Wang, "DeepID : Deep Learning for Face Recognition Machine Learning with Big Data."

[9] Y. Taigman, M. A. Ranzato, T. Aviv, and M. Park, "DeepFace : Closing the Gap to Human-Level Performance in Face Verification."

[10] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07–12–June, pp. 815–823, 2015.

[11] J. Liu, "Targeting Ultimate Accuracy : Face Recognition via Deep Embedding," pp. 1–5.

[12] I. Kemelmacher-Shlizerman, S. Seitz, D. Miller, and E. Brossard, "The MegaFace Benchmark: 1 Million Faces for Recognition at Scale," 2015.

[13] S. Duffner, "Face Image Analysis With Convolutional Neural Networks," *Thesis*, p. 191, 2007.

[14] J.-C. Chen, R. Ranjan, A. Kumar, C.-H. Chen, V. M. Patel, and R. Chellappa, "An End-to-End System for Unconstrained Face Verification with Deep Convolutional Neural Networks," *2015 IEEE Int. Conf. Comput. Vis. Work.*, pp. 360–368, 2015.

[15] L. Huang, Y. Yang, Y. Deng, and Y. Yu, "DenseBox: Unifying Landmark Localization with End to End Object Detection," pp. 1–13, 2015.

[16] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proc. 2001 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition. CVPR 2001*, vol. 1, p. I-511-I-518, 2001.

[17] W. Wang, J. Yang, J. Xiao, S. Li, and D. Zhou, "Face Recognition Based on Deep Learning," vol. 8, no. 10, pp. 812–820, 2015.

[18] Y. Sun, X. Wang, and X. Tang, "Deep Learning Face Representation by Joint Identification-Verification," pp. 1–9, 2014.

[19] Y. Sun, X. Wang, and X. Tang, "Deeply learned face representations are sparse, selective, and robust," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07–12–June, pp. 2892–2900, 2015.

[20] X. Tang, "DeepID3: Face Recognition with Very Deep Neural Networks," pp. 2–6.