ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FACE RECOGNITION: LITERATURE REVIEW WITH EMPHASIS ON DEEP LEARNING

[1,2]RAJESHWAR MOGHEKAR, [2]SACHIN AHUJA

[1]Hyderabad Institute of Technology and Management, Department of Computer Science and Engineering, Hyderabad, India

[2]Department of Computer Science and Engineering,
Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India

E-mail: [1]rajeshwarm.cse@hitam.org ,[2] sachin.ahuja@chitkara.edu.in

ABSTRACT

Under the broad umbrella of object recognition, Face recognition is one area with active research for last few decades mainly due to its applications and the challenges in the environment where they are used. The face recognition as a biometric authentication can work without much cooperation of human. The current face recognition techniques perform well in constrained environment but performance degrades in unconstrained environment as the images captured may vary in resolution, illumination, pose, occlusion and expressions. This sometimes makes intra class variance more than the inter class variance and leads to misclassification. In this article, an overview of some of the strategies adopted by the researchers to overcome the challenges like pose variation, low resolution and occlusion in unconstrained environment has been discussed. Moreover, this paper reviews the use of deep learning in face recognition to achieve accuracy at par with humans in image classification tasks. We also discuss the challenges with Deep learning and the strategies adopted to overcome them. This paper provide an up-to-date review of face recognition techniques.

Keywords: *Face recognition, Deep Learning, Low resolution, Occlusion, Pose invariant.*

1. INTRODUCTION

Biometric identification is the technique of automatically verifying or identifying a person by behavioral and/or physiological characteristics. Biometric authentication like Face recognition where we compare the image captured with the record in the database to identify an individual, finds applications in Surveillance, Access control, , Law enforcement, Cross border security, Multimedia etc.

As a part of identity science, face biometrics has the benefit of being non-intrusive and passive compared with other biometric modalities such as fingerprint and IRIS [1]. Face recognition is considered to be one of the most successful application on which image analysis is performed [2]. Face recognition technique can be operated either as (i) Face verification also known as Face authentication where the Face is matched against the template face images whose identity is to be checked against and (ii) Face identification or Face recognition where the query face is matched with all the Face images stored in the database. [3].

Face recognition system broadly has four modules namely

i. Face Detection/Tracking (in the case of videos)
ii. Face Alignment
iii. Feature Extraction
iv. Comparison of the test face image with the face image in the database.

These modules of face recognition system are depicted in Figure 1. As illustrated, the recognition process starts with identifying the face region in the given image. Tracking is performed to locate a human face if the input is a video sequence. One of the widely used Face detection methods is Viola-Jones algorithm that uses Ada Boost technique.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Further from the detected frame, face part is detected. The detected face region is aligned and adjusted as a major aspect of pre-processing. Features are extracted from the aligned face image using any of the feature extraction techniques and further matched with the features of face images from the gallery.

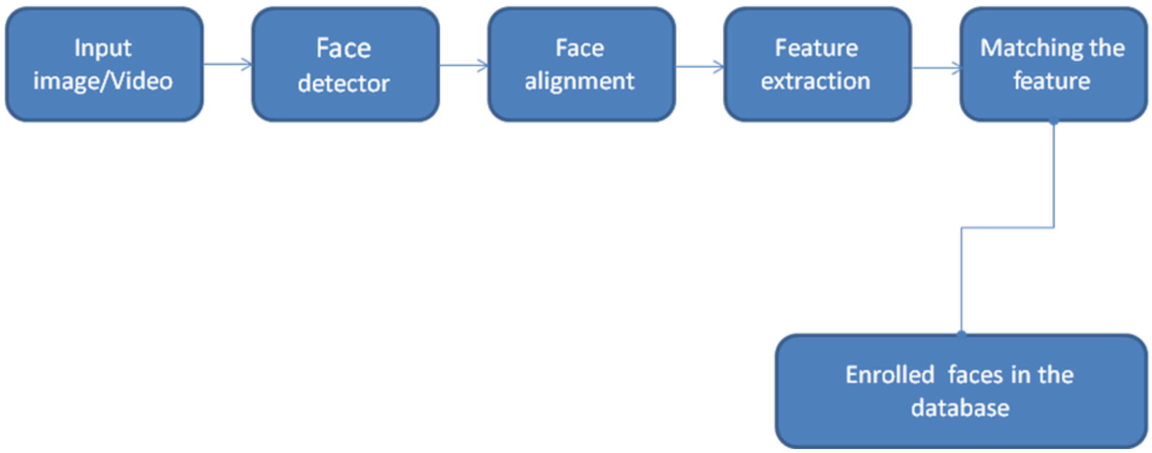


Figure 1: Processing Diagram for Face Recognition.

1.1 Classification of Face Recognition Algorithms

In the last few decades many face recognition methods have been proposed by researchers with different background, which lead to vast and diverse literature. Due to this diverse view in a single system of face recognition it becomes difficult to classify based on the techniques employed. Face recognition techniques have been classified into three categories based on the information from physiological studies as follows [3].

1.1.1 Holistic methods

In these methods, whole face is given as input to a recognition module. The face image is represented in a lower dimension using principal component analysis (PCA) without losing much information, and then reconstructing it [4]. The Eigen pictures are determined from the correlation matrix. These Eigen pictures are the optimal set to represent any picture. Face recognition system was developed based on Eigen pictures [5]. Other techniques like Independent component analysis which is a generalization of principal component analysis[6] and Linear Discriminate Analysis which retrieves vectors to discriminate the classes by increasing the between-class differences, reducing the within-class ones [7] also use whole face image as input.

1.1.2 Feature based methods

The locations and local statistics of the local features on face such as eyes, mouth and nose are extracted and fed into a structural classifier. Kanade in 1974 developed Face recognition system which extracts the local features of the face and defined a face model based on the position, size and the relation between the features [8]. Wiskott et.al use Gabor wavelet transform to represent the local features [9]. Lawrence had proposed a face recognition system which uses local image sample representation, Self Organizing Map (SOM), and Convolution neural network (CNN) [10]. Ahonen et al. divided the Face image into sub regions and calculate the Linear Binary Pattern (LBP) histograms. Later these LBP histograms are combined into global histogram [11]. Liao et al introduced a variation of LBP where they calculate the average values of block sub regions instead of individual pixel [12].

1.1.3. Hybrid Methods

These methods are based on the human vision system which considers both local features and whole face. Pentand in 1994 used both Eigen faces and Eigen modules like Eigen mouth, Eigen eyes, and Eigen nose [13]. Huang et al. in 2003 has used face region and components [14].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2. FACE RECOGNITION IN UNCONTROLLED ENVIORNMENT

The increased usage of surveillance cameras in public areas, ranging from small scale stand-alone camera applications in banks and supermarkets to multiple network close circuit television (CCTV) used for law enforcement in public streets, where there is a need for face recognition technology. Cameras are normally installed to maximize the viewing area. Face recognition system should identify faces captured in uncontrolled environments such as images captured in unpredictable lighting conditions, different angles and distances from the camera, and where the subjects are non-cooperative. We discuss the various challenges of face recognition in videos and the work done by the researchers to reduce their impact on the recognition system.

2.1 Challenges of Face Recognition in Videos

Face recognition is good at applications like surveillance and verification system for access control. Applying the face recognition system at various applications, the outcome of the recognition result is not always accurate and hence this biometric suffer from various challenges. The major challenges with respect to face recognition in videos are described below.

2.1.1 Pose variation

Pose variation while capturing frames from video is extremely normal and this property is a troublesome errand in recognizing faces. Figure 2 displays faces with different poses. As the complete structure of the face is not visible when faces are at different poses, recognition is a difficult task. Abate et al. in 2007 carried work on a survey of two and three dimensional face images. This work expresses that pose changes in face recognition influences largely on the recognition process [15].

The face detection using Viola Jones face detector followed by which they detect 68 landmarks of the face using Dlib shape detector is discussed by Fontaine et al. in their proposed work. They construct Delaunay triangulation mesh by placing equidistant points on the face border and detected landmarks. They perform pose alignment by applying affine transformation on the input image wrap mesh [16]. Ding et al presents the difficulties in Pose Invariant Face recognition and review of existing techniques [17]. Ding et.al retrieve Multi-Directional Multi-Level Dual-Cross Patterns (MDML-DCPs) from face images. MDML-DCPs encodes the invariant characteristics of a face image into patterns that are robust to variations in faces belonging to same class and have high discrimination to faces that belong to different classes [18]. Ganguly et al. performed face recognition of the 3d face images in unconstrained environment with variations in pose, occlusion and lighting. The availability of additional information in the form of depth data in 3D face images is used to rotate. They used Energy Range Face Image model to normalize in terms of the pose variation and occlusion restoration [19].

zhu et al. in 2015 presented facial recognition algorithm by morphing the input images to the model. A High-Fidelity Pose and Expression Normalization (HPEN) method is developed with 3D Morphing Model (3DMM) to normalize expression and pose in the image to frontal face. The model follows two steps, Firstly a land marking marching assumption is done to detect several feature points of face on the given image. Secondly, the whole image is mesh into 3D object to eliminate pose and expression variances using identity preserving 3D transformation [20]. In 2013 Ho et al. used variant of the Belief propagation (BP) Algorithm and Markov Random Fields (MRFs) to generate a frontal view from a face image with pose. The given probe image is first classified as frontal or non-frontal view using SVM. The non-frontal probe image is split into grids with overlapping patches. The patches from frontal view are created from the global optimal set of local warps. The frontal face thus generated is used for the face recognition [21]. In 2011 shroff et. al. develop a face similarity measure which is invariant to pose, expression and illumination. The similarity between the probe image with the images in the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 2: Faces with Varying Pose, (Honda/UCSD dataset).

library is used to create the ordered list. Thus each face image has its own ordered list which is referred as signature of the face. To determine the similarity of probe image its signature is compared with signatures of the images in the library. They used the data driven approach using Doppelganger list for each image from the library. The similarity of the pair can be determined by comparing the similarity of the list computed with the doppelganger list [22].

2.1.2 Occlusion

Occlusion is one of the major drawbacks that affects recognition rate in face recognition technique. Occlusion can occur due to facial accessories or any other object covering your face as the users need not cooperate with the camera when the faces are captured. As the part of the face is covered, the recognition rate reduces drastically when a system is built. Figure 3 shows the occlusions on a face that may occur due to the usage of accessory like a sun glass. This will deteriorate the recognition rate of a face. Wright et al in 2009 stated that lot of investigation has been done in the projection from high to low dimension. But the problem is which features to use. Recently with compressed sensing the feature space is no longer critical, what is important is that the dimension of the feature space is sufficiently large and that the sparse representation is correctly computed. This method outperformed other techniques for corrupted and occluded images [23].
Rui Min et al. in 2011 worked on improving recognition rate of faces that are occluded by facial accessories. In this work, authors have considered sunglasses and scarf as the facial accessories. They developed a method where the presence of sunglass and scarf are detected using Gabor wavelets, Support Vector Machines (SVM) and Principle Component Analysis (PCA). Once face with occlusion of accessories is detected, the recognition is performed from the non-occluded region. This is done using block-based local binary patterns [24].
Li et al. in 2013 proposed a morphological graph model that describes the morphological structure of the occlusion. Incorporating the errors in occluded part and non-occluded part, authors proposed structured sparse error coding for face recognition from occlusion [25]. Alyuz et al. in 2013 developed a 3-D face recognition system that is robust to occlusions. Missing data is handled using subspace analysis technique. Non occluded patches are utilized for construction [26].

2.1.3 Low resolution

Face images captured from surveillance video at longer distance have low resolution. When the person is not close to the camera, the face region will have very few pixels as shown in figure 4. These low resolution images have less detail which may lead to misclassification [27].

Figure 3: Faces with occlusion due to Sun glass, (AR dataset).

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 4: A typical frame from surveillance video (from CIVIAR database).

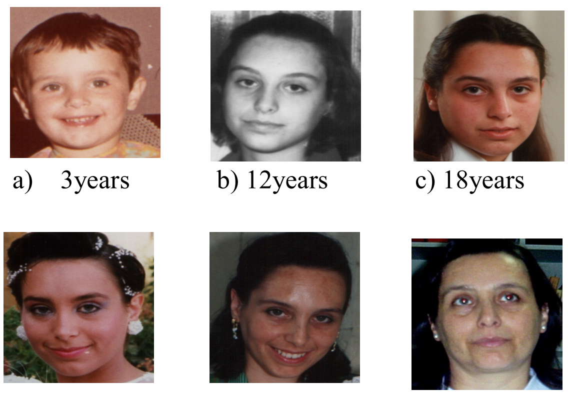
Studies show that for existing algorithms, minimum resolution of Face image lies between 32x32 and 64x64 [28]. Zou et al. in 2012 stated that the recognition performance of the existing methods on VLR face image will degrade dramatically. As most of the image details of VLR face image are lost and it contains very minimal information. They categorize the algorithms into two approaches namely, example based and maximum a posterior (MAP) based. They developed the relationship between the High Resolution image and the VLR image, later apply the relationship to recover the High resolution images from VLR. Face images with frontal view are used in the experiments [29].
Super-resolution (SR) is a method to construct a high-resolution (HR) image from its low-resolution (LR) image [30]. Theoretically, applying SR technique on VLR face image, the reconstructed HR image can be used for face recognition. Gunturk et al. applied MAP-based SR method to reconstruct the Eigen face coefficients for face recognition [31]. Wang et al. employed example-based approach to generate HR images for face recognition [32].

Biswas et al. extract SIFT descriptors from HR gallery and LR probe images and transform them to a space in which inter-Euclidean distances approximates with the distances calculated for all the descriptors using HR frontal images. Multidimensional scaling is used to learn the desired transformation. They use SIFT based descriptors as the input feature which are represented for the fiducial locations of the face image. Tensor analysis based approach is used to predict rough locations of the facial landmarks and approximate pose. The scale factor between the HR gallery (60x55) and LR probe images is fixed at 3. Multiple Biometric Grand Challenge (MBGC) and FRGC dataset are used. The recognition performance using the proposed approach significantly improves over the SIFT combined with PCA features [33].

3. DEEP LEARNING

In last few years the attention of the researchers towards deep learning has increased dramatically due to the results achieved. It finds applications in various fields including speech recognition, audio recognition, computer vision, machine translation and drug design, where the results are on par to and in few cases outperform human experts. Deep learning has been there for few decades but its usage has tremendously increased in last few years mainly due to availability of large datasets and fast computing GPU's. The top results in competition Imagenet Large Scale Visual Recognition Challenge (ILSVRC) which is image classification task with 1000 different classes clearly depicts that Convolution Neural Networks(CNN) provide better results. Convolution Neural Networks has an advantage that it requires less weights or parameters when compared with fully connected layers as the weights are being shared. Convolution Neural Networks have a significant impact in the field of computer vision. In 2012 during ILSVRC competition Alexnet achieved 15.3% top-5 error rate [34]. In majority of the CNN models, Softmax is used as the loss function. Wen et al. introduced center loss to improve the discrimination in the features learned from the deep neural network. They combined Softmax and center loss function with the goal to reduce intra class variation and increase inter class variation. The results achieved on Labeled Face in the Wild (LFW), Mega Face challenge and You Tube Faces are on par with state of the art on face verification and recognition tasks [35].

One of the major challenges in face recognition is that with the variation in age of a person, the intra class variation has drastic impact on the performance of a model developed. Figure 5 show face images of a particular subject with variations in age, i.e. a,b,c,d,e and f are the images of the same subject at age 3,12,18,23,31 and 38 years respectively. Wen et al. proposed a frame which learns the face features invariant to age using a Convolutional Neural Network. They extract age invariant features using latent factor fully connected network layer from the CNN features. They carried out experiments on FGNET and CACD-VS datasets, the results obtained are on par with the state of the art [36].



a) 3years b) 12years c) 18years

d) 23years e) 31years f) 38years

Figure 5: Face images of a subject from FGNET dataset.

Wang et al. proposed a framework to search the probe image against gallery of images which combines CNN and commercial of the shelf (COTS). They compare the deep features of the probe image extracted from the CNN with deep features of the face images in the gallery and prepare top-k similarities. Later they rank it once again by combining the similarity scores of deep features and COTS. They observed that their performance improved when deep features were combined with COTS face matcher [37].

In 2015 Liu et al cascaded two CNN's one for face detection and the other for attributes representation. The first CNN is pre-trained with Image Net dataset which consists of 1000 different classes. Later they train the model with 2000 images with single faces and 2000 background images from sun dataset. They calculate the sum of scores over all the 500 Windows designed using Edge box and normalized to Window size. The second CNN has four convolution layers, where C1 and C2 share the filter globally and the convolution layers C3 and C4 filters share locally. Their model outperforms face detectors Face++, SURF Cascade, DPM and ACF multi view. The attributes predicted using it outperforms PANDA-W by 10% on celeb A dataset [38]. In 2014 Sun et al. used deep learning model that extracts the Deep hidden Identity features (Deep ID) which are high level features. The deep network consists of four convolution layers with max pooling, a gully connected layer and softmax output layer. The face images were aligned Using two eye centers and the two corners of the mouth that were detected using facial point detection. They trained the network on CelebFaces that has total 87,628 face images of 5436 different celebrities collected from the Internet. To know the impact of training on large dataset they used CelebFaces+, which contains 202,599 face images of 10,177 celebrities. The Face verification accuracy of 97.45% is achieved on LFW [39].

In face recognition we need to extract or represent the features in way which reduces the intra personal variations and increases inter personal variations. Sun et al in 2014 have represented the features of a face in way which reduces the intra personal variations and increases inter personal variations. In this paper they designed a deep convolution network which extract DEEPID2 features from different identities which increases inter personal variations and the DEEPID2 features from same identities reduces the intra personal variations which is essential for face recognition [40]. Zhu et al. in 2013 designed a deep network with layers that extract the features and the layers for reconstruction. Feature extraction layers encode face images into FIP features and reconstruction layers transform them into images in the canonical view. Face images with variations in pose and illumination are given as input to this network which reconstructs a frontal pose with neutral illumination (canonical view) of the target identity. Results indicate there is significant improvement of the existing methods, when those are applied on the reconstructed face images. Experiments were conducted on MultiPIE face database with 337 identities and 754,204 total images. They captured face images with 15 different poses and 20 different illuminations for each identity. They applied PCA, SC, LDA, and PCA+LDA directly on pixels of the original images and the reconstructed images, they observe that each of the above methods can be improved at least 30% on average [41].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.1 CHALLENGES IN DEEP LEARNING

3.1.1 Large datasets

It might be difficult to get large datasets for all applications. To overcome this problem either they use data augmentation or data synthesis. In 2018, Hu. et al. used data synthesis to create large datasets from a smaller set. This is helpful for deep neural networks which require large number of images for training. They used just 10000 images to generate 1.5 million images using data synthesis. This is different from data augmentation where they perform cropping, translation and scaling without affecting the semantics. In cases like near to Infrared images it is difficult to get large number of images. This technique can be used to synthesize images which are required to train on deep neural networks. The results indicate that the model trained on the synthesized data performs on par with those trained on large datasets [42]. Taigman et al in 2014 trained the deep neural network with nine layers on Social Face Classification dataset which contains 4.4 million labeled faces of 4030 people each with 800 to 1200 faces, with 5% of recent faces were used for testing. They have applied 3d – alignment and frontalization, validated the architecture on LFW dataset and achieved mean accuracy recognition of 97.5% [43].

3.1.2 Over fitting

The model gives high accuracy in the training but performance degrades on test data. Deep neural networks may encounter over fitting which may solved using dropout. Generally dropout refers to eliminating some neurons in hidden layer to overcome over fitting due to the noise in the input. Dropout is applied by eliminating both input and output connections of a neuron permanently from hidden layer. Dropout is not applied during the final softmax layer and during testing phase. Dropout helps for the network to be more generalized. Srivastava et. al. in 2014 used dropout to solve the over fitting problem [44]. They perform on CIFAR-10, IMAGENET, and MNIST datasets and found that dropout can be applied on any classification problem.

3.1.3 Computational resources

To train the model with large data it requires GPU's and consumes lot of time. An alternate solution could be to apply transfer learning. Traditionally machine learning and deep learning algorithms were developed to solve specific tasks and work in isolation. The concept of utilizing the learning from task to solve the new related task is referred as transfer learning. Figure 6 depicts the traditional system where the learning systems work in isolation. Figure 7 depicts how knowledge in terms of features, weights etc. learnt on one task is transferred to other new task.

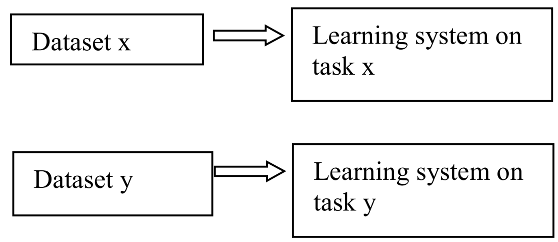


Figure 6: Traditional isolated single task learning

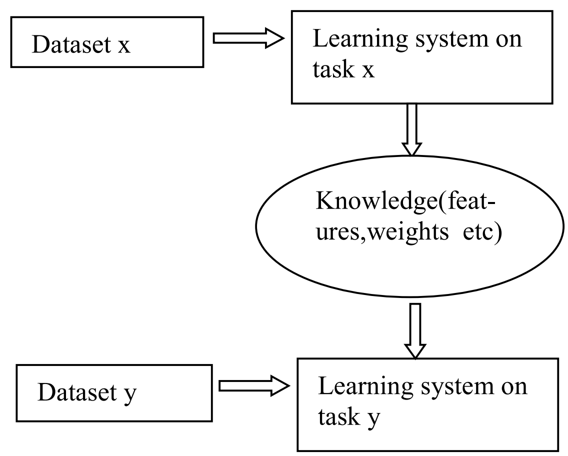


Figure 7: Transfer Learning

Transfer learning can be implemented by extracting the features from the convolution layer preceding the final layer of a pre-trained model and train the classifier on top of it. In this method we do not update the weights of pre-trained model. The other way would be to replace the final layer and update the weights of few layers using back propagation. This way fine tuning the pre-trained is adopted when we have enough data for the new task.

Wang Ya. et al. in 2017 used parameters of a VGG face model instead of training their model with random weights. They handle the difficulties of Face Recognition in Real-world Surveillance Videos with Deep Learning Method by an approach called FINE-TUNING. They employ a variety of schemes to improve the efficiency such as collecting and forming a large dataset of real-world surveillance by detecting the faces and labeling using an autonomous algorithm. There after the collected dataset is purified using a GRAPH CLUSTERING approach for the elimination of redundancy of collected data and any in appropriate labeling in the data. Results show a great variation of the accuracy between the pre trained VGG face model and the fine-tuned VGG model as 83.6% to 91.4% i.e. it shows a margin of 7.8% on the accuracy of the model before and after fine-tuning the final layers of the model [45].

Sun et al. in 2015 introduced two deeper models, one model has nine convolutional layers and the second model has four convolutional layers and five inception layers. The deep networks used in DEEPID2 are shallower when compared to the DEEPID3. Motivated by the Supervisory signals used in Deep Learning Face Representation by Joint Identification- Verification they were added to final layer and few inception layers [46].

3.1.4 Hyperparameters

Hyperparameters have a significant impact on the performance of a deep learning model. These variables provide information about structure of the model and how a network would be trained. We can improve the performance of model by changing values of the hyperparameters with help of validation test. A lot of research is going on to get the right hyperparameters for deep learning model using some algorithm instead of manual selection which may not provide the performance in a stipulated time. Hyperparameters like number of layers, number of hidden units in a layer, dropout, activation function are linked with structure of the network where as learning rate, momentum, batch size, number of epochs are linked with training of network. Dahl et al. in 2013 used Bayesian optimization technique instead of traditional manual tuning to obtain improved deep neural network. The results based on the experiments carried out, suggest that Relus and dropout provide the synergy to achieve word error rate improvements compared to the state of art [47]. Dohman et al. in 2015 presented technique to speed up the search of hyperparameters automatically for deep neural networks using a learning curve model which terminates the training by extrapolating the performance based on few steps if it is expected to give low performance [48]. Bregstra et al. in 2012 proposed Random search as an alternate for grid search and manual which are generally used for hyperparamater optimization. The results obtained depict that random search provides better model than pure grid search with a less computational effort [49]. Bregstra et al. in 2013 performed hyperparameter optimization using Tree of Parzen Estimators (TPE). Their results show that they achieved best configuration than random search on LFW and PUBFIG83 datasets [50]. Agostinelli et al. in 2014 proposed activation function which is adaptive in nature with each neuron learning a linear activation function using gradient descent. They improve performance over deep neural networks using fixed Relu by using their adaptive activation function on CIFAR-10 and CIFAR-100 dataset [51]. Ramachandran et al. in 2017 introduced an activation function by name swish which performs better than Relu on Imagenet using deeper models [52]. Sutskever et al. in 2013 demonstrated the impact of initialization and momentum on deep learning models [53]. Smith et al. in 2017 claim that increasing the batch size during training would result into the same performance as what we get by decaying the learning rate with an added advantage of less number of parameter updates [54]. Simonyan et al. had shown how the depth of the convolution neural network impacts the accuracy. They observed that by having 16-19 layers they could achieve better results compared with Alexnet.
Apart from that the model generalizes on different tasks or datasets [55]. Iandola et al. introduced Squeezenet which performs on par or better than Alexnet with 50 times less number of parameters. They adopted few steps to achieve the compressed network like using 1x1 filters instead of 3x3 filters, reducing the number of input channels using squeeze layers and down sample at the later part of the network to achieve high accuracy [56]. In [34] they observed the impact of activation function Relu and Tanh with Convolution Neural Network having four layers. Figure 6 shows that they achieved error rate of 25% using Relu indicated by solid line much faster than when used Tanh activation indicated by dotted lines.

3.1.5 Summary and Discussions

Different techniques used to overcome the challenges encountered in unconstrained environment to recognize the face have been discussed in the first part of the paper. The authors have proposed methods to overcome a particular challenge in unconstrained environment like pose variation or low resolution or occlusion. In order to use face recognition for authentication in real time applications the system should handle all the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

problems simultaneously. In the latter part of the paper we discussed how Deep learning has impacted the face recognition along with the challenges to implement face recognition using Deep learning. We even discuss alternatives suggested by researchers to overcome the problems. The observation made after surveying the research literature is that we require a system which recognizes face with different variations and with less number of training images.

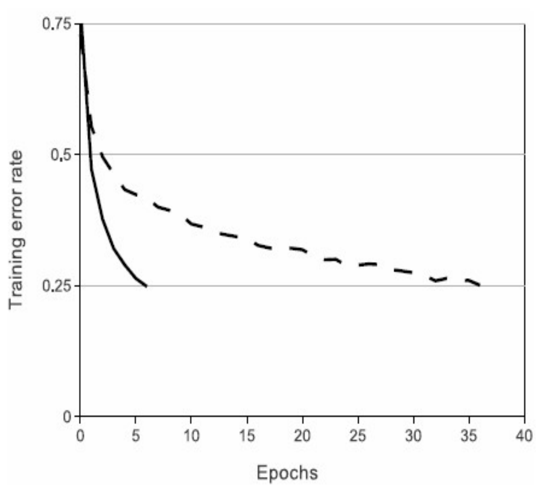


Figure 6: Four Layered convolution network trained CIFAR-10 dataset. Relu indicated by solid line and Tanh by dotted lines [34]

4. CONCLUSION AND FURTHER DIRECTION

Face recognition as a biometric has advantage over other biometrics that it requires little or no cooperation from human. We have presented the work carried out by the researchers to overcome the challenges like pose variation, occlusion and low resolution which makes face recognition difficult in real time applications. As deep learning has results on par with human in image classification tasks, we have presented researchers contribution towards face recognition using deep learning. There are many challenges in deep learning like dataset size, computational power and hyperparamters which affect the performance. From the literature review we feel there is scope to develop a deep learning model which gives accurate results in unconstrained environment with minimum number of parameter updates, hyperparameter optimization and to explore how the transfer learning can be effectively utilized to make the recognition system efficient.

REFRENCES:

[1] Jain, A. K., Ross, A., & Prabhakar, S. (2004). An introduction to biometric recognition. *IEEE Transactions on circuits and systems for video technology*, *14*(1).

[2] Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM computing surveys (CSUR)*, *35*(4), 399-458.

[3] Suganya, S., & Menaka, D. (2014). Performance evaluation of face recognition algorithms. *International Journal on Recent and Innovation Trends in Computing and Communication*, *2*(1), 135-140.

[4] Sirovich, L., & Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *Josa a*, *4*(3), 519-524.

[5] Turk, M. A., & Pentland, A. P. (1991, June). Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on* (pp. 586-591). IEEE.

[6] Bartlett, M. S., Movellan, J. R., & Sejnowski, T. J. (2002). Face recognition by independent component analysis. *IEEE Transactions on neural networks*, *13*(6), 1450-1464.

[7] Yu, H., & Yang, J. (2001). A direct LDA algorithm for high-dimensional data—with application to face recognition. *Pattern recognition*, *34*(10), 2067-2070.

[8] Kanade, T. (1974). Picture processing system by computer complex and recognition of human faces.

[9] Wiskott, L., Krüger, N., Kuiger, N., & Von Der Malsburg, C. (1997). Face recognition by elastic bunch graph matching. *IEEE Transactions on pattern analysis and machine intelligence*, *19*(7), 775-779.

[10] Lawrence, S., Giles, C. L., Tsoi, A. C., & Back, A. D. (1997). Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks*, *8*(1), 98-113.

[11] Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE transactions on pattern analysis and machine intelligence*, *28*(12), 2037-2041.

[12] Liao, S., Zhu, X., Lei, Z., Zhang, L., & Li, S. Z. (2007, August). Learning multi-scale block local binary patterns for face recognition. In

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

International Conference on Biometrics (pp. 828-837). Springer, Berlin, Heidelberg.

[13] Pentland, A., Moghaddam, B., & Starner, T. (1994). View-based and modular eigenspaces for face recognition.

[14] Huang, J., Heisele, B., & Blanz, V. (2003, June). Component-based face recognition with 3D morphable models. In *International Conference on Audio-and Video-Based Biometric Person Authentication* (pp. 27-34). Springer, Berlin, Heidelberg.

[15] Abate, A. F., Nappi, M., Riccio, D., & Sabatino, G. (2007). 2D and 3D face recognition: A survey. *Pattern recognition letters*, *28*(14), 1885-1906.

[16] Fontaine, X., Achanta, R., & Süsstrunk, S. (2017). Face Recognition in Real-world Images. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (No. EPFL-CONF-224338).

[17] Ding, C., & Tao, D. (2016). A comprehensive survey on pose-invariant face recognition. *ACM Transactions on intelligent systems and technology (TIST)*, *7*(3), 37.

[18] Ding, C., Choi, J., Tao, D., & Davis, L. S. (2016). Multi-directional multi-level dual-cross patterns for robust face recognition. *IEEE transactions on pattern analysis and machine intelligence*, *38*(3), 518-531.

[19] Ganguly, S., Bhattacharjee, D., & Nasipuri, M. (2015). Illumination, pose and occlusion invariant face recognition from range images using ERFI model. International Journal of System Dynamics Applications (IJSDA), 4(2), 1-20.

[20] Zhu, X., Lei, Z., Yan, J., Yi, D., & Li, S. Z. (2015). High-fidelity pose and expression normalization for face recognition in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 787-796).

[21] Ho, H. T., & Chellappa, R. (2013). Pose-invariant face recognition using markov random fields. *IEEE transactions on image processing*, *22*(4), 1573-1584.

[22] Schroff, F., Treibitz, T., Kriegman, D., & Belongie, S. (2011, November). Pose, illumination and expression invariant pairwise face-similarity measure via doppelgänger list comparison. In *Computer Vision (ICCV), 2011 IEEE International Conference on* (pp. 2494-2501). IEEE.

[23] Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., & Ma, Y. (2009). Robust face recognition via sparse representation. *IEEE transactions on pattern analysis and machine intelligence*, *31*(2), 210-227.

[24] Min, R., Hadid, A., & Dugelay, J. L. (2011, March). Improving the recognition of faces occluded by facial accessories. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on* (pp. 442-447). IEEE.

[25] Li, X. X., Dai, D. Q., Zhang, X. F., & Ren, C. X. (2013). Structured sparse error coding for face recognition with occlusion. *IEEE transactions on image processing*, *22*(5), 1889-1900.

[26] Alyuz, N., Gokberk, B., & Akarun, L. (2013). 3-D face recognition under occlusion using masked projection. *IEEE Transactions on Information Forensics and Security*, *8*(5), 789-802.

[27] Zou, W. W., & Yuen, P. C. (2012). Very low resolution face recognition problem. *IEEE Transactions on Image Processing*, *21*(1), 327-340.

[28] [28] Lui, Y. M., Bolme, D., Draper, B. A., Beveridge, J. R., Givens, G., & Phillips, P. J. (2009, September). A meta-analysis of face recognition covariates. In *Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on* (pp. 1-8). IEEE.

[29] Zou, W. W., & Yuen, P. C. (2012). Very low resolution face recognition problem. *IEEE Transactions on Image Processing*, *21*(1), 327-340.

[30] Park, S. C., Park, M. K., & Kang, M. G. (2003). Super-resolution image reconstruction: a technical overview. *IEEE signal processing magazine*, *20*(3), 21-36.

[31] Gunturk, B. K., Batur, A. U., Altunbasak, Y., Hayes, M. H., & Mersereau, R. M. (2003). Eigenface-domain super-resolution for face recognition. *IEEE transactions on image processing*, *12*(5), 597-606.

[32] Wang, X., & Tang, X. (2003, June). Face hallucination and recognition. In *International Conference on Audio-and Video-Based Biometric Person Authentication* (pp. 486-494). Springer, Berlin, Heidelberg.

[33] Biswas, S., Aggarwal, G., Flynn, P. J., & Bowyer, K. W. (2013). Pose-robust recognition of low-resolution face images. *IEEE transactions on pattern analysis and machine intelligence*, *35*(12), 3037-3049.

[34] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in*

neural information processing systems (pp. 1097-1105).

[35] Wen, Y., Zhang, K., Li, Z., & Qiao, Y. (2016, October). A discriminative feature learning approach for deep face recognition. In *European conference on computer vision* (pp. 499-515). Springer, Cham.

[36] Wen, Y., Li, Z., & Qiao, Y. (2016). Latent factor guided convolutional neural networks for age-invariant face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4893-4901).

[37] Wang, D., Otto, C., & Jain, A. K. (2017). Face search at scale. *IEEE transactions on pattern analysis and machine intelligence*, *39*(6), 1122-1136.

[38] Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep learning face attributes in the wild. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 3730-3738).

[39] Sun, Y., Wang, X., & Tang, X. (2014). Deep learning face representation from predicting 10,000 classes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1891-1898).

[40] Sun, Y., Chen, Y., Wang, X., & Tang, X. (2014). Deep learning face representation by joint identification-verification. In *Advances in neural information processing systems* (pp. 1988-1996).

[41] Zhu, Z., Luo, P., Wang, X., & Tang, X. (2013). Deep learning identity-preserving face space. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 113-120).

[42] Hu, G., Peng, X., Yang, Y., Hospedales, T. M., & Verbeek, J. (2018). Frankenstein: Learning deep face representations using small data. *IEEE Transactions on Image Processing*, *27*(1), 293-303.

[43] Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). 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).

[44] Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *Journal of machine learning research*, *15*(1), 1929-1958.

[45] Wang, Y., Bao, T., Ding, C., & Zhu, M. (2017, June). Face recognition in real-world surveillance videos with deep learning method. In *Image, Vision and Computing (ICIVC), 2017 2nd International Conference on* (pp. 239-243). IEEE.

[46] Sun, Y., Liang, D., Wang, X., & Tang, X. (2015). Deepid3: Face recognition with very deep neural networks. *arXiv preprint arXiv:1502.00873*.

[47] Dahl, G. E., Sainath, T. N., & Hinton, G. E. (2013, May). Improving deep neural networks for LVCSR using rectified linear units and dropout. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 8609-8613). IEEE.

[48] Domhan, T., Springenberg, J. T., & Hutter, F. (2015, June). Speeding up automatic hyperparameter optimization of deep neural networks by extrapolation of learning curves. In *Twenty-Fourth International Joint Conference on Artificial Intelligence.*

[49] Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research,* 13(Feb), 281-305.

[50] J. Bergstra, D. Yamins, and D. D. Cox, ''Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures,'' in Proc. Int. Conf. Mach. Learn., 2013, pp. 115–123..

[51] Agostinelli, F., Hoffman, M., Sadowski, P., & Baldi, P. (2014). Learning activation functions to improve Deep Neural Networks *arXiv preprint arXiv:1412.6830*.

[52] Ramachandran, P., Zoph, B., & Le, Q. V. (2017). Searching for activation functions. *arXiv preprint arXiv:1710.05941*.

[53] Sutskever, I., Martens, J., Dahl, G., & Hinton, G. (2013, February). On the importance of initialization and momentum in deep learning. In *International conference on machine learning* (pp. 1139-1147).

[54] Smith, S. L., Kindermans, P. J., Ying, C., & Le, Q. V. (2017). Don't decay the learning rate, increase the batch size. arXiv preprint arXiv:1711.00489.

[55] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[56] Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. *arXiv preprint arXiv:1602.07360*.