An Efficient Scheme of Deep Convolution Neural Network for Multi View Face Detection

Shivkaran Ravidas
Dept. of Electrical Engineering, School of Engineering, Gautam Buddha University, Gautam Buddha Nagar, INDIA
E-mail: mailmekaran@gmail.com

M.A. Ansari
Dept. of Electrical Engineering, School of Engineering, Gautam Buddha University, Gautam Buddha Nagar, INDIA
E-mail: mahmadiitr@gmail.com

Abstract—The aim of this paper is to detect multi-view faces using deep convolutional neural network (DCNN). Multi-view face detection is a challenging issue due to wide changes in appearance under different pose expression and illumination conditions. To address challenges, we designed a deep learning scheme with different network structures to enhance the multi view faces. More specifically, we design cascade architecture on convolutional neural networks (CNNs) which quickly reject non-face regions. Implementation, detection and retrieval of faces will be obtained with the help of direct visual matching technology. Further, a probabilistic calculation of resemblance among the images of face will be conducted on the basis of the Bayesian analysis for achieving detection of various faces. Experiment detects faces with ±90 degree out of plane rotations. Fine-tuned AlexNet is used to detect multi view faces. For this work, we extracted examples of training from AFLW (Annotated Facial Landmarks in the Wild) dataset that involve 21K images with 24K annotations of the face.

Index Terms—Face detection, multi view face detection, deep learning, convolutional neural network (CNN) and Computer vision.

I. Introduction

Multi-view detection for the face is very challenging when it is viewed from the fixed view; therefore, it is significant to adopt multi-view faces. Multi-view face detection system is used to predict the upright faces in images with 90 degrees ROP (rotation out of plane) pose changes. The meaning of rotation invariant is to predict faces with 360 degrees RIP (rotation in plane) pose variations. The multi-view face detection system produces rotated windows of an image and their integral windows of an image for every classifier which perform operations of parallel classification to predict non-frontal and non-upright faces in images. Multi-view detection of a face can be detected by building few detectors, all consequent to a particular view. Detection of the face was one of the main technologies for enabling natural interaction between human and computer [1]. The performance of systems for recognizing the face relies extremely on representing the face that is physically coupled with most of the variations in the face type like expression, view, and illumination. As images of a face are mostly noticed in unique views, the main threat is to unpick the identity of face and representations of the view. Large efforts are contributed for extracting features for identity by hand like Gabor proposed by Liu et al. [2], LBP (Local Binary Pattern) by Ahonen et al. [3], and SIFT (scale invariant feature transform) by D. Lowe [4]. The best practice of detecting the face obtains the above features on the images of face landmarks with various scales and concatenates them into feature vectors at high dimension as explained by Simonyan et al. [5, 6]. CNN (convolutional neural networks) was involved in the community of computer vision by the storm, effectively enhancing the art state in most of the applications. Main significant elements for the success of CNN methods are the accessibility of high amount of training data as illustrated by Simonyan and Lin et al. [7]. CNNs are neural networks which are hierarchical those layers in convolution exchange with subsampling layers, suggestive of complex and simple cells in the fundamental visual cortex. Even though neural networks are adapted to tasks of computer vision for obtaining good performance for generalization, it is good to add before knowledge into architecture of the network. CNN aims to adopt spatial information between images pixel and thus they are on the basis of discrete convolution. According to Li et al. [8] detection of a face is formulated as an issue for classification for separating patterns for the face from non-face patterns. From the perspective of statistics, there are three drawbacks for such issue patterns dimensionality is high usually; probable quantity of non-face patterns is huge and their distribution is not regular. Hence, it is complex to model the distribution of probability in patterns of the face, particularly the patterns in the multi-view face with a unimodal function for density. Issues concern with rotation in profound and thus able to identify faces across various views are not simple. Most of the investigators and researchers

addressed such issue by constructing multiple views on the basis on detecting the face (multi-view face detection) that is to categorize the sphere of view into certain small segments and to build one detector on all segments [3, 9].

Eigenface is an approach adopted for recognition of the face. Due to the algorithm's simplicity, deployment of an eigenface system for recognition becomes easy. It is effective in processing, storage and time. The accuracy of this approach relies on various factors. As eigenface considers the value of the pixel for comparing the projection, the accuracy would minimize with the differing intensity of the light. Face recognition can be improved by using hybrid approach that is combining more than one technique [10]. Image preprocessing is needed for achieving a satisfactory outcome. The benefit of such algorithm is that they were developed particularly for those aim what makes the image system very effective. A shortcoming of eigenface is sensitive to conditions of lightening and head position. Identifying the Eigen values and eigenvectors is time consuming [11].

The paper is organized as follows. In section 2, related work is described. Section 3 describes face detection method. Here, mainly multi view face detection and deep convolutional neural network is discussed. Section 4, describes the implementation method in detail. Section 5 contains the experimental result and discussion. Finally, in section 6, the conclusions are drawn.

II. RELATED WORK

Girshick et al. [12] has proposed CNN based detection method called as R-CNN which has achieved stated of the art result on VOC 2012. R-CNN follows the "regions based" approached for detection and recognition. Zhang et al. [13] and Park et al. [14] utilized the multiresolution approach for object detection. Neural network is being used historically for face detection [15, 16]. Yang et al. [17] and Mahmoodi et al. [18] provides detailed face detection survey. Vaillant have applied neural networks for detecting faces in uncluttered images. They designed a convolutional neural network that can be trained to detect the presence or absence of a face in a given image. This will scan the whole image at all possible locations.

Rowley et al. [16] has developed a neural network for upright frontal face detection. Later in 1998 [19] the method was extended for pose invariant face detection. Neural networks are adopted in most of the applications such as issues in recognition of pattern, recognition of character, recognition of the object and autonomous robot driving. The major purpose of this network in the recognition of face is the training feasibility of the system for capturing the difficult class of patterns in the face. Zhang et al. [20] studied about enhancing multi-view detection of a face with multi-task deep CNN. Multi-view detection of the face is the main issue because of dramatic appearance modifications under different pose, expression conditions or illumination. This work adopted multi-task deep CNN for building a post filter in order to enhance the accuracy of multi-view detection of the face. The concept is to learn the non-face or face decision

together with estimation of facial pose and localization for the facial landmark. This study achieved the performance of state of the art on the challenging data set of face detection. Thus, it is inferred from the study that developed method assist in learning the non-face or face decision together with estimation of facial pose and localization for the facial landmarks. Farfade et al. [21] conducted a research to examine multi-view detection of the face using deep CNN. Developed framework does not need landmark or pose annotation and can identify faces in a large choice of orientations with the help of a single model. DDFD (Deep Dense Face Detector) is not dependent on common modules in deep objects for learning the methods for detection like bounding-box regression and segmentation of an image. We compared the developed method with R-CNN and few other methods for detecting the face which is designed particularly for multi-view detection of the face, for example, DPM-based and cascade-based. It was demonstrated that developed detector method can achieve better or similar outcomes even without adopting information or pose annotation about facial landmarks. In addition to these, this work examined the performance of the developed method on different images of the face and identified that there is a link between the distribution of positive illustration in the set for training and scores of the developed method for detection. Thus, it is clear that developed detector method can achieve better or similar outcomes even without adopting information or pose annotation about facial landmarks.

According to Parkhi et al. [22], the recognition of face from either a set of faces or single photograph tracked in a video. Two major contributions were made in this particular research. First and foremost, we have developed a procedure which can assemble a wide range of dataset, with the small noise of label while reducing the quantity of manual annotation included. One of the main concept was to adopt weaker classifiers for ranking the data given to the annotators. At the same time, it was noted that such procedure was designed for faces however appropriate for other classes of objects and fine-grained responsibilities. The second contribution was to demonstrate that deep CNN, with appropriate training and without any additions can produce outcomes when compared with state of the art. Thus, it can be concluded that deep CNN can outperform well without any additions and appropriate training than other counterparts as well as reduce the quantity of manual annotation.

Li et al. [23] analyzed about CNN cascade for detecting the face. Developed detector estimates the image as input at low resolution to refuse non-face regions and cautiously process the difficult region at higher resolution for exact identification or detection. Nets for calibration are brought in the cascade for accelerating identification and enhance the quality of bounding box. Sharing the benefits of CNN, developed detector for the face is robust to large variations in the visual image. Apart from these, it was noted that on the public FDDB (face detection data set and benchmark) developed detector performs well as compare to the state-

of-the-art methods. It was also pointed out that developed detector is fast, achieve 14 frames per second for typical video graphics array images on the central processing unit and can be accelerated to 100 frames per second on the graphical processing unit. Thus, it was clear from the findings of the research are sharing the benefits of CNN, developed detector for the face is robust to large variations in the visual image.

According to the study by Zhu et al. [24] analyzed multi-view perception (MVP) through the deep model for learning the identity of face and view representations. This work developed a generative deep network known as MVP to mimic the capable of perception at multi-view in the primate brain. MVP can disentangle the representations of view and identities are obtained as input for the image and also create a full views spectrum of the image as input. From the findings of the experiment, it was demonstrated that detection features of MVP achieve better outcome and performance on recognition of face than counterparts like state-of-the-art methods. In addition to these, it was demonstrated that modeling the factor for view representation as a continuous variable allows MVP for predicting and interpolating images beneath the viewpoints that are unobserved in data for training, which imitate the reasoning human capacity. Thus, it can be inferred from the analysis that detection features of MVP achieve better outcome and performance on recognition of face than counterparts like state-of-the-art methods.

III. FACE DETECTION

A. Multi View Face Detection

We can define Face detection as the process of extracting faces from the given images. Hence, the system should positively identify a certain region as a face. According to Yang et al. [17], Rath et al. [25] and Erik Hjelmas et al. [26], face detection is a process of finding regions of the input image where the faces are present. A lot of work has been done in detecting faces in still and frontal faces in plane as well as complex background [27]. With the rapid advancement in the field of information technology and computational power, computers are more interactive with humans. This human computer interface (HCI) is done mostly via traditional devices like mouse, keyboard, and display. One of the most important mediums is the face and facial expression. Face detection is the first step in any face recognition system. Detecting face is well studied problem in the computer vision [28, 29]. Contemporary detectors of the face can effortlessly identify near front faces. Complexities in detecting the face come from two aspects such as large space for searching of probable face sizes, positions and large visual differences of human faces in a chaotic environment. Former one imposes a requirement for the efficiency of time while latter one needs a detector for face to perfectly addressing a binary issue in classification.

It was noted that uncontrolled issue in detecting face are extreme illuminations and exaggerated expressions can lead to large differences in visual in the appearance of the face and affect the face detector robustness. This is significant to develop a method to properly detecting the faces as pointed out in [4, 7]. Therefore, this particular work intends to concentrate on detecting the face with the help of multi-view face using deep convolution neural network.

B. Deep Convolutional Neural Network (DCNN)

Convolutional neural network (CNN) are very popular in the field of computer vision. One of the reasons is availably of large amount of training data. Deep convolution CNN is not only used for face detection but also for face alignment [30]. For obtaining the best performance of such method, it has to highly tune number of nodes, layers, rates for learning and so on [31]. The drawback in the approach of a neural network is that when the quantity of classes maximizes. In template matching, other templates for the face are exploited from various prospects for characterizing single face. Such algorithms are not cost effective and cannot be easily carried out as stated in [32].

The main characteristic of Deep Learning is that it is capable of making abstractions, by building complex concepts from simpler ones. Given an image, it is capable of learning concepts such as cars, cats or humans by combining sets of basic features, such as corners or edges. This process is done through successive "layers" that increase the complexity of the learned concepts. The idea of depth in Deep Learning comes precisely from these abstraction levels.

IV. IMPLEMENTATION METHOD

In the implementation, detection of face and retrieval of the image will be attained with the help of direct visual matching technology. A probabilistic computation of resemblance among the images of the face will be conducted on the basis of the Bayesian analysis for achieving various detection of the face. After this, a neural network will be developed and trained in order to enhance the outcome of Bayesian analysis.

Next, to that, training and verification will be adopted to test other images which involve similar face features. Deep learning can be performed by supervisory signals given by Sun et al. [33].

$$\text{Ident}(f, \text{t}, \emptyset_{id}) = -\int \sum_{i=1}^{n} \log \hat{p}_i = -\log \hat{p}_t \quad (1)$$

Where f is the feature vector, t represents target class and $\emptyset_{id}$ is softmax layer parameter, p_i is the target probability distribution ($p_i = 0$ for all i except $p_t = 1$), $\hat{p}_i = 1$ is the predicted probability distribution. The verification signal regularize feature and reduces intra personal variations given in [33].

$$Verif(f_i, f_j, y_{ij}, \emptyset_{ve}) =$$
$$\begin{cases} \frac{1}{2}\|f_i - f_j\|_2^2 & if \ y_{ij} = 1 \\ \frac{1}{2}\max(0, m - (\|f_i - f_j\|_2)^2 & if \ y_{ij} = -1 \end{cases} \quad (2)$$

$$Verif(f_i, f_j, y_{ij}, \emptyset_{ve}) = \frac{1}{2}\left(y_{ij} - \sigma(wd + b)\right)^2 \quad (3)$$

Where $\emptyset_{ve} = \{w, b\}$; are denote shifting parameters and learning scaling, σ represented as sigmoid function and y_{ij} is denoted as binary target of two compared facial images relate to same identity.

Further operation of convolution is represented as:

$$y^{j(r)} = \max(0, b^{j(r)} + \sum_i k^{ij(r)} * x^{i(r)}) \quad (4)$$

Where x^i is input map and y^j is output map, k^{ij} is convolution between input and output.

Maxpooling is given by:

$$y_{j,k}^i = \max_{0<m,n<s}\{x_{j.s+m,k.s+n}^i\} \quad (5)$$

Where output map pools over $s \times s$ non-overlapping region.

$$y_j = max\left(0, \sum_i x_i^1 . w_{i,j}^1 + \sum x_i^2 . w_{i,j}^i + b_j\right) \quad (6)$$

Where x^1, w^1, x^2, w^2 represent the neurons and weights in 3rd and 4th convolutional layers. Output of ConvNet is n-way software to predict the distribution of probability over n-unique identities [34].

$$y_n = \frac{\exp(y_i')}{\sum_{j=1}^n \exp(y_j')} \quad (7)$$

DCNN is mostly adopted for classification and also adopted for detection and recognizing the face. Most of them consider the cascade strategy as well as consider batches with various locations and scales as inputs. At the same time, it was also noted that to a vast amount, these operations maximize time and space difficulty. Other than conventional methods, DCNN does not need to initialize locations' shape. Therefore, we can neglect getting jammed in local optima for avoiding the poor initialization of shape.

A. CNN Structure

The CNN structure which is adopted in the present study consists of 12-net CNN, 12-calibration-net, 24-net, 24-calibration-net, 48-net.

12-net CNN

It is the first CNN that scans or tests the image quickly in the test pipeline. An image having the dimensions of $w * h$ having the pixel spacing of 4 with 12x12 detection windows for such type of image 12-net CNN is suitable to apply. This would result a map of:

$$\left(\left(\frac{W-12}{4} + 1\right) \times \left(\frac{H-12}{4} + 1\right)\right) \quad (8)$$

Point on the image map defines detection window of 12x12 onto the testing image. The minimum size of the face acceptable for testing an image is 'T'. Firstly, an image pyramid is built through the test image in order to cover the face from varied scales. At each level an image pyramid is created, it is resized by 12/T which would serve as an input image for 12-net CNN. Under this structure 2500 detection windows are created as shown in Figure 1.
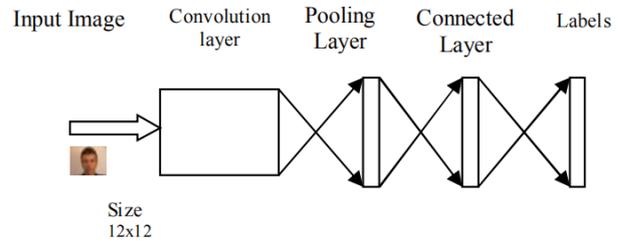

Fig.1. 12-net CNN

12-Calibration-net

For bounding box calibration, 12-calibration-net is used. Under this the dimension of the detection window is (x, y, w,) where 'x' and 'y' are the axis, 'x' and 'h' are the width and height respectively. The calibration pattern adjusts itself according to the size of the window is:

$$(x - \frac{x_n w}{s_n}, y - \frac{y_n h}{s_n}, \frac{w}{s_n}, \frac{h}{s_n}) \quad (9)$$

In the present study number of patterns i.e. N=45. Such that:

$$s_n \in \{0.87, 0.95, 1.2, 1.13, 1.25\}$$

$$x_n \in \{-0.19, 0, 0.19\}$$

$$y_n \in \{-0.19, 0, 0.19\}$$

The image is cropped according to the size of detection window that is 12*12 which would serve as an input image to 12-calibration-net. Under this CNN average result of the patterns are taken because the patterns obtained as an output are not orthogonal. A threshold value is taken i.e. t in order to remove the patterns which are not the confidence patterns.

24-net CNN

In order to further lower down the number of the detection windows used, a binary classification of CNN called 24-net CNN is used. The detection window which remained under the 12-calibration net are taken and then resized to 24*24 image and then this image is re-evaluated using 24-net. Also under this CNN, multi-resolution structure is adopted, through this the overall overhead of the 12-net CNN structure got reduced and hence the structure becomes discriminative. As shown in Figure 2.

48-calibration-net CNN

It is the last stage or sub-structure of CNN. The number of calibration patterns used is same as in case of 12-calibration-net i.e. N=45. In order to have more accurate calibration, pooling layer is used under this CNN sub-structure.
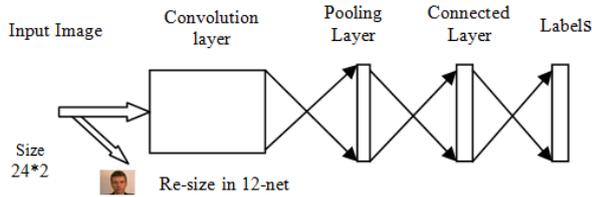

Fig.2. 24-net CNN

48- net CNN

It is the most effective CNN used after 24-calibration-net but is quite slower. It follows the same procedure as in 24-net. This procedure used in this CNN is very complicated as compared to rest of the CNN sub structures. It also adopts the multi resolution technique as in case of 24-net.

B. Proposed Algorithm for DCNN

This particular work develops an algorithm for detecting the face using multi-view with the help of deep convolution neural network. The steps of implementation are described below:

Step 1: In the implementation, detection of face and retrieval of image will be attained with the help of direct *visual matching technology* which match the face directly. This technology makes use of similarity metrics of an image which can either be normalized correlation or it can be Euclidean distance, which corresponds to the approach called *template matching*. The similarity between the two images is measured through *similarity measure*, denoted by $S(I_a, I_b)$,Where, I_a and I_b are the two images between which the similarity is being measured.

Step 2: The next step is measuring probabilistic similarity or Δ(the measure of intensity difference between the two images) given by Probabilistic similarity or $\Delta= (I_a - I_b)$. This calculation of resemblance among the images of face will be conducted on the basis on the Bayesian analysis for achieving various detection of face.

Step 3: The probabilistic calculation of resemblance also supports multiple face detection. In order to characterize the various types of image variations were used for statistical analysis. Under this the similarity measure S (I_a, I_b) between the pair of images I_a and I_b is given in terms of posteriori probability (interpersonal variation) is provided by:

$$S(I_a, I_b) = P(\Omega_I)P(\Omega_I|\Delta)/\{P(\Omega_I)P(\Omega_I|\Delta) \\ +P(\Omega_E)P(\Omega_E) \qquad (10)$$

If the multi-view face detection is done for a single person then $P(\Omega_I|\Delta) > P(\Omega_E|\Delta)$ or it can be said that $S(I_a , I_b) > \frac{1}{2}$.

Step 4: Further a neural network will be developed and trained in order to enhance the outcome from Bayesian analysis.

Step 5: Next to that, training and verification will be adopted to test other images which involve similar face features.

Implementation of the code is done step by step as follows:

1. *First, the DCNN object is created.*
2. *Second after this Graphical user interface is initialized.*
3. *Then MCR (Misclassification rate) calculation is initialized and plot of MCR id created defining the current epoch, iteration, RMSE (Root Mean Square Error), MCR value of the image data.*
4. *Training data is being loaded.*
5. *Training data is preprocessed, errors are deleted and then image data is simulated.*
6. *After the simulation the multi-faces are detected in the image shown in the red rectangular boxes*
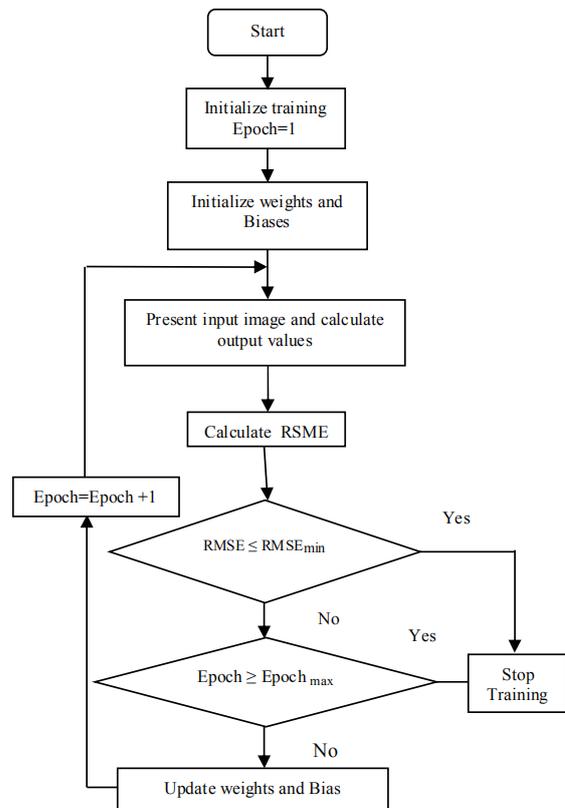

Fig.3. Flow chart of training Process

The flowchart for CNN training progress is shown in the Figure 3. The plot of RMSE (Root Mean Square Error) and MCR (Misclassification Rate) in training is shown Figure 4. The screenshot depicts the DCNN training progress is also shown in Figure 4.

The below equation is the CNN which is trained to minimize the risk of soft max loss function.

$$R = \sum_{x_i \in \beta} \log | prob(x_i|y_i) | \qquad (11)$$

Here 'β' represents the batch used in iteration of stochastic gradient decent and label is $'x_i'$ and $'y_i'$. Hessian calculation progress is started. Current epoch used for this is 3.00. Iteration value used for this research is 759.00. RMSE value used for this research is 0.18. MCR value used for this research is 0.90. Here 'theta' used is $8.000e^{-05}$. Plot of RMSE in training is showed in zig zag lines. Plot of MCR in training is showed in curved lines.
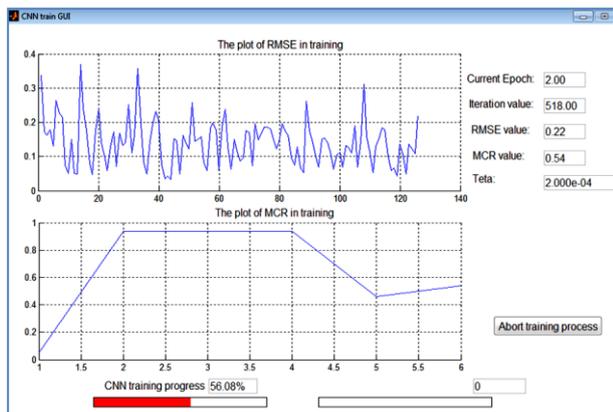

Fig.4. DCNN training Process

V. RESULT AND DISCUSSION

Detection of face is formulated as a categorization issue to isolate patterns of face from non-face patterns. There are many reasons for this issue such as patterns dimensionality is high. It is complicated to develop model the possibility distribution of patterns in the face, particularly the multi-view patterns for face with a unimodal function for density and probable amount of non-face pattern is enormous and their distribution is not regular. To detect face across various views is very challenging when seeing from fixed view as the face appearance is unique from various views. Method for detecting the multi-view face is to develop a single detector that focus on all face views. Multi-view detection of face can be detecting by building few detectors, all consequent to a particular view. Further, it was stated that in execution-time, if one or more detectors provide positive result for specific sample, then face will be recognized. Multi-view face detection is a challenging issue due to wide changes in appearance under different pose expression and illumination conditions. The modern face detection solutions performance on multi view face set of data is unsatisfactory.

The result of the study is determined on the basis of following parameters both in the presence and absence of multi resolution in proposed CNN structure.

Detection Rate: It is defined as the rate at which the face of a person in an image is detected.

Number of False Detection: The number of the face which are not detected at all, or detected falsely.

It was observed that, in the presence of multiresolution in CNN, which is shown in Figure 5, number of false detections comes to halt (at the 10000 number of falsely detected faces) and the face is detected or the detection rate is achieved.
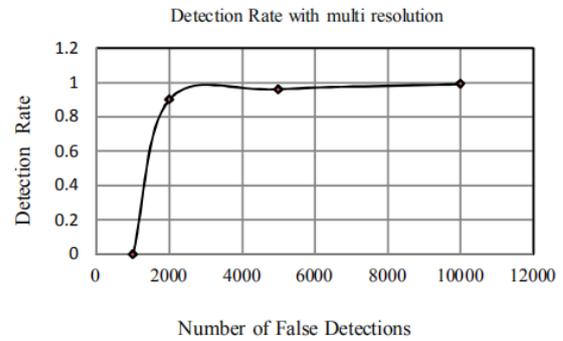

Fig.5. Detection rate with multi resolution in 24-net CNN

However, without the use of multi-resolution in CNN, a greater number of faces are detected falsely as compared to that of multi-resolution shown in Figure 6. This work develops an algorithm for detecting the face using multi-view with the help of deep convolution neural network. The main concept of the algorithm is influencing the high ability of DCNN to classify and extract the feature for learning the single classifier for detecting faces from different views and reduce the computational difficulty to simplify the detector architecture.
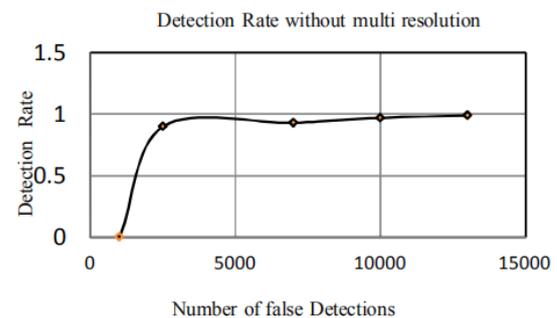

Fig.6. Detection rate without multi resolution in 24-net CNN

FDDB (Face Detection data Set and benchmark) dataset [35] contains annotated faces. This is a large-scale face detection benchmark. It uses ellipse faces annotation and also defines two types of evaluations. One is discontinuous score evaluation and other is continuous score evaluation. To augment the data, we randomly flip the illustration of training. In the fine-tuned deep network, it is probable to take either approach of sliding or region-based window for obtaining the final detector for face. For this particular work, we have chosen a sliding approach of window since it has less difficulty and it is not dependent of additional modules like choosy search.

Fig.7. Detection results: (a) Original Image given for detection (b) Image at preprocessing stage (c) Detected face position with CNN.
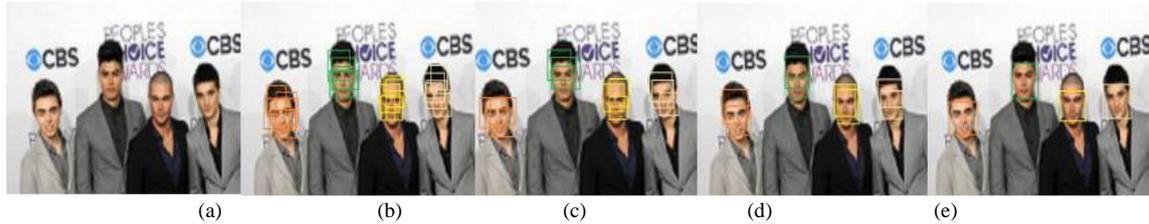

Fig.8. Detection results for different CNN structure: (a) Input/Test Image, (b) Image after 12-net CNN, (c) Image after 24-net CNN, (d) Image after 48-net CNN, (e) Output face detected Image

In this work, classifier of face, when compared to AlexNet involve 8 layers in which first five layers are convolutional and then final three layers are completely connected. For this work, we first transformed the completely linked layers into the convolutional one to reshape the parameters of the layer. In our cascaded CNNs, we have used AlexNet [36] to apply ReLU nonlinearity function after pooling layer and a fully connected layer.
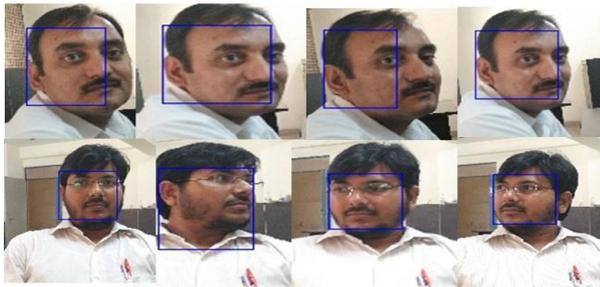

Fig.9. Pose invariant Face detected Images

This will be probable to efficiently execute the CNN on any size images and uses a heat map in classifying the face. Every point in the heat map indicates the response of CNN, possibility of involving a face, for its consequent 227*227 region in real image. To recognize the face of various sizes, investigator scaled the images up and down, and acquired new heat maps. Here, we have attempted various schemes for scaling and identified that image for rescaling three times per octave provides good result. In addition to these, to detect the face are enhanced by adopting bounding-box module for regression. The overall test sequence is shown in Figure 7. The detail explanation for all the CNNs has been already given in previous sections. First of all, test image is applied to the system, a 12-net structure will scan the whole image and quickly rejects about 90% of detection windows.

Remaining detected window will be processed by 24 calibrated CNNs. In next subsequent stages, the highly overlapped window will be removed. Then a 48 net will take detected windows and evaluate the window with calibrated boundary box and produces as output as detected boundary box. Figure 8 shows all detection stages with different stages. Examples of the input images for two different identities with generated multi view output results are illustrated in Figure 9. In this Figure, detected face for the various angle and poses for left and right profile faces including frontal face are shown. Our detector gives results for images with varying poses with resolution.

As per image processing is concern, CNNs provides lots of advantages as compared to fully connected and unconstrained neural network architectures. Typical images are large and without a specialized architecture, hence it is difficult to manage when presented to the network. This problem resolves by neural network by using preprocessing of the images. In general CNN architecture is more suitable for such application as compare to conventional neural network.

VI. Conclusion

In this work, we have presented deep CNN cascade structure which produces fast detection by rejecting non-face regions in varying resolutions for accurate detection. The convolutional structure used is AlexNet to detect the face. For this research we extracted examples of training from AFLW dataset that involve 21K images with 24K annotations of the face. We randomly sampled images sub-windows and adopted them as positive illustration if it was higher than a 50 per cent intersection over union. It was noted that proposed method performs well in terms of accuracy and the detection rate. Developed method easily identifies the face and produces the better result in

the fastest time. Effectiveness of the developed method is compared and Contrasted with existing methods and techniques. It was noted that proposed method performs well in terms of accuracy and the detection rate. Exploiting the power of CNN, given method work well in images with large variations. In future this work can be extended to better strategies for sampling and more techniques can be adopted to enhance the detection through augmentation of data to further enhance the effectiveness of the developed method to detect the round, occluded and rotated faces.

REFERENCES

[1]   YI, Fang, L. I. Hao, and J. I. N. Xiaojie, "Improved Classification Methods for Brain Computer Interface System", *International Journal Computer Network and Information Security*, ISSN: 2074-9104, Vol. 4, Issue: 2, pp.15-21, Mar-2012.

[2]   Liu, Chengjun, and Harry Wechsler, "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition", *IEEE Transactions on Image processing,* ISSN: 1057-7149, Vol. 11, Issue:4, pp. 467-476, Apr-2002

[3]   Ahonen, T., Hadid, A. and Pietikainen, M., "Face Description with Local Binary Patterns: Application to Face Recognition", *IEEE Transactions on Pattern Analysis & Machine Intelligence*, ISSN: 0162-8828, Vol. 12, Issue: 12, pp. 2037-2041, Dec-2006

[4]   Lowe, David G., "Distinctive Image Features from Scale-Invariant Keypoints", *International Journal of computer Vision*, ISSN: 1573-1405, Vol. 60, Issue: 0.2, pp.91-110, Nov-2004.

[5]   Simonyan, K., Parkhi, O. M., Vedaldi, A., & Zisserman, A., "Fisher Vector Faces in the Wild", *In British machine Vision Conference (BMVC)*, ISBN: 1-901725-49-9, Vol. 2, No. 3, pp. 4, Sep-2013.

[6]   Simonyan, Karen, Andrea Vedaldi, and Andrew Zisserman, "Learning Local Feature Descriptors using Convex Optimisation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, ISSN: 0162-8828, Vol. 36, Issue: 8, pp.1573-1585, Aug-2014

[7]   Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D. and Zitnick, C. L., "Microsoft coco: Common Objects in Context", *European conference on Computer Vision,* ISBN: 978-3-319-10602-1, Springer, Cham, pp. 740-755, Sep- 2014.

[8]   Li, Haoxiang, Zhe Lin, Jonathan Brandt, Xiaohui Shen, and Gang Hua, "Efficient Boosted Exemplar-Based Face Detection", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, ISSN: 1063-6919, pp. 1843-1850, 2014.

[9]   Matsumoto Yoshio and Alexander Zelinsky, "An Algorithm for Real-Time Stereo Vision Implementation of Head Pose and Gaze Direction Measurement*", Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition,* ISBN: 0-7695-0580-5, pp. 499, Mar-2000.

[10]   Ansari, M. A., and Aishwarya Agnihotri, "An Efficient Face Recognition System Based on PCA and Extended Biogeography-Based Optimization Technique", *Indian Journal of Industrial and Applied Mathematics,* ISSN: 1945-919X, Vol. 7, Isuue: 2, pp. 285-305, 2016.

[11]   Jaiswal, Sushma, "Comparison Between Face Recognition Algorithm-Eigenfaces, Fisherfaces and Elastic Bunch Graph Matching", *Journal of Global Research in Computer Science*, ISSN: 2229-371X, Vol. 2, Issue: 7, pp.187-193, Aug-2011.

[12]   R. Girshick, J. Donahue, T. Darrell, and J. Malik., "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, ISBN: 978-1-5386-0733-6, pp. 580-587, 2017.

[13]   W. Zhang, G. Zelinsky, and D. Samaras, "Real-time Accurate Object Detection using Multiple Resolutions", *Proceeding IEEE International Conference on Computer Vision*, ISBN: 9781424416301, pp. 1-8, Oct-2007.

[14]   D. Park, D. Ramanan, and C. Fowlkes, "Multiresolution Models for Object Detection", *European Conference on Computer Vision,* ISSN: 1611-3349 *Springer*, pp. 241-254, Sep-2010.

[15]   Vaillant, Régis, Christophe Monrocq, and Yann Le Cun, "Original Approach for the Localisation of Objects in Images", *IEE Proceedings Vision, Image and Signal Processing,* ISSN: 1359-7108, Vol. 141, Issue: 4, pp.245-250, Aug-1994.

[16]   Rowley, Henry A., Shumeet Baluja, and Takeo Kanade, "Neural Network-Based Face Detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, ISSN: 0162-8828, Vol. 20, Issue: 1, pp.23-38, Jan-1998.

[17]   Yang, Ming-Hsuan, David J. Kriegman, and Narendra Ahuja, "Detecting Faces in Images: A Survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* ISSN: 0162-8828, *Vol.* 24, Issue:1, pp. 34-58, Jan -2002

[18]   Mahmoodi, Mohammad Reza, and Sayed Masoud Sayedi, "A Comprehensive Survey on Human Skin Detection", *International Journal of Image Graphics and Signal Processing,* ISSN: 2074-9082, *Vol.* 8, issue: 5, pp. 1-35, May-2016.

[19]   Rowley, H., Baluja, S. and Kanade, T., "Rotation Invariant Neural Network-Based Face Detection", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, ISSN: 1063-6919, 1998, pp. 38-44, Jun-1998.

[20]   Zhang, C. and Zhang, Z., "Improving Multi-View Face Detection with Multi-Task Deep Convolutional Neural Network", *Applications of Computer Vision (WACV), IEEE Winter Conference,* ISBN: 9781479949847 pp.1036-1041, Mar-2014.

[21]   Farfade, S.S., Saberian, M.J., and Li, L.J.,"Multi-view Face Detection using Deep Convolutional Neural Networks", *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, ISBN: 978-1-4503-3274-3 ACM, pp. 643-650, Jun- 2015.

[22]   Parkhi, O.M., Vedaldi, A. and Zisserman, A., "Deep Face Recognition", *British Machine Vision Conference (BMVC),* ISBN: 1-901725-53-7, Vol. 1, Issue: 3, p.6, Sep-2015.

[23]   Li. Haoxiang, Zhe Lin, Xiaohui Shen, Jonathan Brandt, and Gang Hua, "A Convolutional Neural Network Cascade for Face Detection", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, ISBN: 9781467369657, pp. 5325-5334, 2015.

[24]   Zhu, Zhenyao, Ping Luo, Xiaogang Wang, and Xiaoou Tang, "Multi-view Perceptron: a Deep Model for Learning Face Identity and View Representations", *Advances in Neural Information Processing Systems*, ISBN: 978-1-5108-0041-0, pp.217-225, 2014.

[25]   Rath, Subrat Kumar, and Siddharth Swarup Rautaray, "A Survey on Face Detection and Recognition Techniques in Different Application Domain", *International Journal of*

Modern Education and Computer Science", ISSN: 2075-017X, Vol. 6, Issue: 8, pp.34, 2014

[26] Hjelmås, Erik, and Boon Kee Low, "Face Detection: A Survey*", Computer Vision and Image Understanding*, ISSN: 1077-3142, Vol. 83, Issue: 3, pp.236-274, Sep-2001.

[27] Sheikh Amanur Rahman, M.A. Ansari and Santosh Kumar Upadhyay, "An Efficient Architecture for Face Detection in Complex Images", *International Journal of Advanced Research in Computer Science and Software Engineering*, ISSN: 2277-128X, Vol. 2, Issue 12, pp. 211-216, Dec-2012.

[28] Mariappan, M., Fang, T.W., Nadarajan, M. and Parimon, N., "Face Detection and Auto Positioning for Robotic Vision System", *International Journal of Image, Graphics and Signal Processing*, ISSN: 2074-9082, Vol. 7, Issue: 12, pp. 1-9, Nov-2015.

[29] Anam, R., Rahman, M., Haque, M.O. and Islam, M.S., "Computer Vision Based Automation System for Detecting Objects", *International Journal of Intelligent Systems and Applications*, ISSN: 2074-9058, Vol. 7, Issue: 12, p.65, Nov- 2015

[30] Sharma, Kartikeya, Shivkaran Ravidas, and M. A. Ansari, "A Novel Technique for Face Alignment Using Deep Convolutional Neural Networks", *Indian Journal of Industrial and Applied Mathematics*, ISSN: 1945-919X, Vol. 8, Issue: 1, pp.107-117, 2017.

[31] Yang, Bin, Junjie Yan, Zhen Lei, and Stan Z. Li, "Aggregate Channel Features for Multi-view Face Detection*", International Joint Conference on Biometrics (IJCB)*, *In Proceeding on IEEE International Joint Conference*, ISBN: 9781479935857, pp. 1-8, Sep-2014.

[32] Jyoti S. Bedre ,Shubhangi Sapkal, "Comparative Study of Face Recognition Techniques: A Review", *Proceeding Published in International Journal of Computer Applications,* ISSN: 0975 - 8887, Vol. 12, 2012

[33] Sun Y, Chen Y, Wang X, Tang X., "Deep Learning Face Representation by Joint Identification-verification", *Advances in Neural Information Processing Systems,* ISBN: 9781510800410, pp. 1988-1996, Dec-2014

[34] Sun, Yi, Xiaogang Wang, and Xiaoou Tang, "Deep Learning Face Representation from Predicting 10,000 Classes", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, ISSN: 1063-6919, pp. 1891-1898, 2014.

[35] Jain. Vidit and Erik Learned-Miller," Fddb: A Benchmark for Face Detection in Unconstrained Setting", *Technical Report UM-CS-2010-009, Vol.88, University of Massachusetts,* Amherst, 2010.

[36] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton., "ImageNet Classification with Deep Convolutional Neural Networks", *Advances in Neural Information Processing Systems*, ISBN: 9781627480031, pp. 1097-1105, 2012.

Authors' Profiles

Shivkaran Ravidas has received B.E. in Electronics Engineering from Dr. Baba Saheb Ambedkar Marathwada University, Aurangabad, India, in 2002 and M.E .in Electronics Engineering from SGGS College of Engg. and Technology, Nanded, India, in 2006. Currently he is working as research scholar in the Department of Electrical Engineering, Gautam Buddha University , Greater Noida, India. He has published several papers in reputed journals & conferences. His main areas of research interest are image processing, machine learning and Multi-view face Detection.

M.A. Ansari received B.Tech. (Electrical Engineering) in 1998 from AMU, Aligarh, India, M.Tech. & Ph.D. (Electrical Engineering) in 2001 & 2009 respectively from Indian Institute of Technology, Roorkee, India. The author is associated with Gautam Buddha University and is currently working in the Department of Electrical Engineering. He has wide national and international experience of teaching and has visited several countries. He has published several papers in reputed national and international journals and conferences. His research interest includes medical image& signal processing, Biomedical Instrumentation, Soft computing and wavelet applications. He is senior member of IEEE and ISIAM.