IJASRM

International Journal of Advanced Scientific Research and Management, Special Issue II, Oct 2018

www.ijasrm.com

ISSN 2455-6378

# Multi View Face Detection using Deep Learning

**Shivkaran Ravidas[1] and M.A. Ansari[2]**

[1,2]Dept. of Electrical Engineering, School of Engineering, Gautam Buddha University, Greater Noida, India

**Abstract**

The real life face images have large visual variation such as pose, expression and lighting. To address these challenges, we propose a cascade structure of convolutional neural network. The aim of this paper is to investigate detection of the face with the help of deep convolution neural network. Implementation, detection and retrieval of faces will be obtained with the help of direct visual matching technology. Further, the probabilistic measure of the similarity of the face images will be done using Bayesian analysis. 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.

***Keywords***—*Convolutional neural network (CNN),. Face detection, Multi- view face detection, Deep learning.*

## 1. 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. The meaning of rotation invariant is to predict faces with 360 degrees RIP (rotation in plane) pose variations. 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. 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. 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. [1, 2]. 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. 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. 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 [1,3].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 [4]. 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 [5].

## 2. Face Detection

Face detection is a computer based technique in which for a given arbitrary image, face detection system determines wither or not there is any faces in

the image and, if there are faces present in the image ,return the location of image and also the extend of each face. There have been thousand of different methods and approaches are reported in the literature. The most prominent and systematic survey are [6] and recent one is in [7]. Face detection is one of the most important subset of computer vision but not limited to face tracking [8], facial analysis[9], face recognition[10], gender and age recognition [11], face relighting and morphing [12], beauty assessment [13], face shape reconstruction [14], image retrieval and digital photo album. Face detection is the integral part of smart phones and digital cameras. Finally, most of the social networking like Facebook uses face detection for tagging.

### *A. Deep Convolutional Neural Network (DCNN)*

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 [15].

Multi-view detection of the face is the main issue because of dramatic appearance modifications under different pose, expression conditions or illumination. 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.

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.

Li et al. [16] 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. 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. [17] 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.

## 3. Implementation Method

The operation of convolution is represented as:

$$y^{j(r)} = \max(0, b^{j(r)} + \sum_i k^{ij(r)} * x^{i(r)} \quad \text{--(1)}$$

Where, $x^i$ is input map and $y^j$ is output map, $k^{ij}$ is convolution between input and output.
Maxpooling is given by:

$$y^i_{j,k} = \max_{0<m,n<s} \{x^i_{j.s+m,k.s+n}\} \quad \text{--(2)}$$

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

$$y_j = \max(0, \sum_i x^1_i . w^1_{i,j} + \sum_i x^2_i . w^2_{i,j} + b_j) \text{--(3)}$$

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 [18].

IJASRM
International Journal of Advanced Scientific Research and Management, Special Issue II, Oct 2018
www.ijasrm.com
ISSN 2455-6378

$$y_i\ y = \frac{\exp(y_i')}{\sum_{j=1}^{n} \exp(y_j')} \quad \text{--(4)}$$

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.

***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.

*1) 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:

$$\left( (\frac{w-12}{4}) + 1) \times (\frac{h-12}{4} + 1) \right) \quad \text{--(5)}$$

At each level an image pyramid is created, it is r esized by 12/T which would serve as an input image for 12-net CNN as shown in Figure 1.

*2) 12-Calibration-net:* For bounding box calibration, 12-calibration-net is used. Under this the dimension of the detection window is $(x, y, w, h)$, where, $'x'$ and $'y'$ are the axis, $'w'$ and $'h'$ are the width and height respectively. The calibration pattern adjusts itself according to the size of the window is:

$$\left( x - \frac{x_n w}{s_n}, y \frac{y_n h}{s_n}, \frac{w}{s_n}, \frac{h}{s_n} \right) \quad \text{--(6)}$$

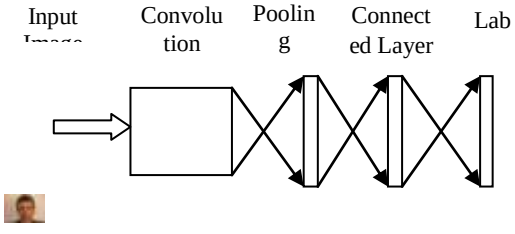



**Figure 1.** 12-net CNN

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.

*3) 24-net CNN:* To lower down number of detection window 24-net CNN is used. Also under this CNN, multi-resolution structure is adopted as shown in Figure 2.

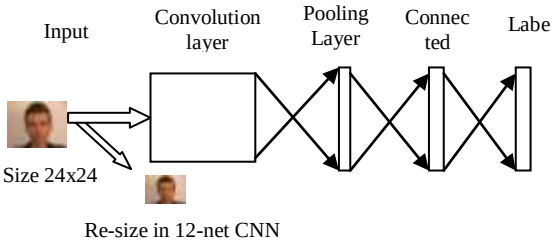
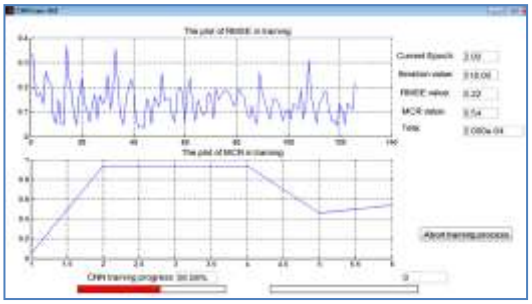



**Figure 2.** 24-net CNN

*1) 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.



The plot of RMSE in training and plot of MCR in training is shown in the above figure. CNN training progress is shown in the Figure 3. The screenshot explains about the DCNN training is also shown in Figure 4.

**Figure 3.** 24-net CNN

## 4. Results and Discussions

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,

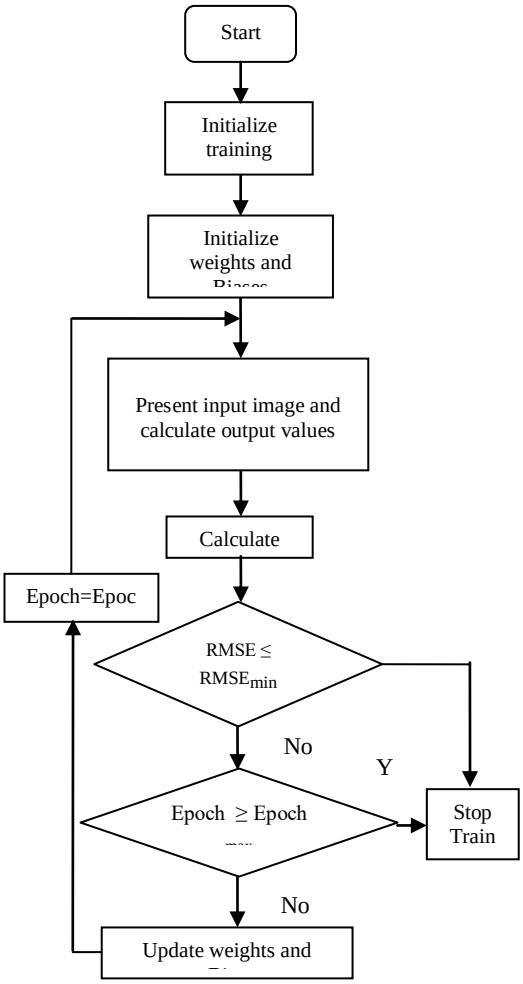



**Figure 4.** Flow chart of training process

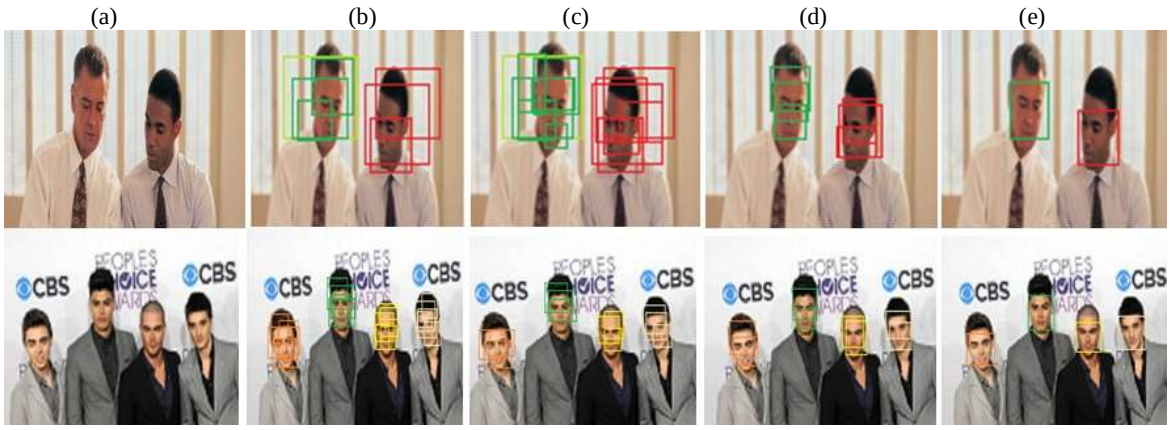



**Figure 5.** Detection results for different CNN structure : (a) Input/Test Image, (b) Image after 12-net CNN , (c) Image after 24-net CNN , (d) Image after 48-net CNN, (e) Output face detected Image

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. Examples of the input images for two different identities with generated multi view output results are illustrated in Figure 8. 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.The result of the study is determined on the basis of detection rate and number of false detection. Under this it was observed that in the presence of multi-resolution in CNN which is shown in Figure 7, number of false detection comes to halt (at the 10000 number of falsely detected faces) and the face is detected or the detection rate is achieved. In the fine tuned deep network, it is probable to take either approaches 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.

IJASRM
International Journal of Advanced Scientific Research and Management, Special Issue II, Oct 2018
www.ijasrm.com
ISSN 2455-6378

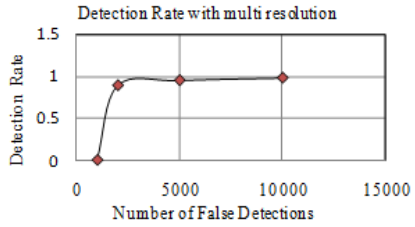



Fig 6. Detection rate with multi resolution in 24-net CNN

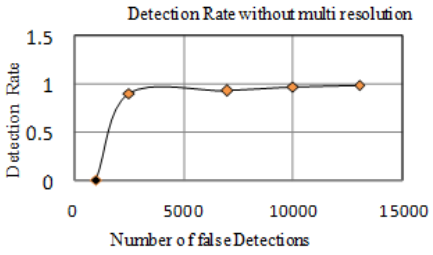



Fig 7. Detection rate without multi resolution in 24-net CNN

**Figure 8.** Pose invariant Face detected Images

FDDB ( Face Detection data Set and benchmark) dataset [19] 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 approaches 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. 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. In our cascaded CNNs, we have used AlexNet [20] to apply ReLU nonlinearity function after pooling layer and a fully connected layer. Examples of the input images for two different identities with generated multi view output results are illustrated in Figure 7. 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.

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 5. 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.

## 5. 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. By, fine-turning AlexNet to detect the face. For this work 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. Developed method easily identifies the face and produces the better results in the fastest time. Effectiveness of the developed method is compared and Contrasted with existing methods and techniques. It observed that proposed method performs better 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.

ISSN 2455-6378

## References

[1] Simonyan, K., Parkhi, O. M., Vedaldi, A., & Zisserman, A. Fisher Vector Faces in the Wild. In *BMVC* (Vol. 2, No. 3, p. 4). (2013, September):

[2] Simonyan, Karen, Andrea Vedaldi, and Andrew Zisserman.: Learning local feature descriptors using convex optimisation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36.8 (2014): 1573-1585.

[3] Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. Microsoft coco: Common objects in context. In *European conference on computer vision* (pp. 740-755). Springer, Cham.(2014)

[4] 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* 7.2 (2016): 285-305

[5] Jaiswal, Sushma. "Comparison between face recognition algorithm-eigenfaces, fisherfaces and elastic bunch graph matching." *Journal of Global Research in Computer Science* 2.7 (2011): 187-193.

[6] Yang, Ming-Hsuan, David J. Kriegman, and Narendra Ahuja.: Detecting faces in images: A survey." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24.1 (2002): 34-58.

[7] 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, Vol. 2,Issue 12, 2012.

[8] Ming-Hsuan Yang, Member, IEEE, David J. Kriegman, Senior Member, IEEE, and Narendra Ahuja, Fellow, IEEE "Detecting Faces in Images: A Survey", IEEE Transaction on pattern analysis and machine intelligence , Vol 24,No.1 January 2002

[9] E. Hjelmas and B. K. Low. Face detection: A survey. *Computer Vision and Image Understanding"*, 83:236– 274, 2001.

[10] S. Zafeiriou, C. Zhang and Z. Zhang,"A survey on face detection in the wild: past, present and future", . Computer Vision and Image Understanding, 138, pp.1-24. 2015

[11] M. Pantic, L. J. M. Rothkrantz, Automatic analysis of facial expressions: The state of the art, Pattern Analysis and Machine Intelligence, IEEE Transactions on 22 (12) (2000) 1424–1445.

[12] W. Zhao, R. Chellappa, P. J. Phillips, A. Rosenfeld, Face recognition: A literature survey, Acm Computing Surveys (CSUR) 35 (4) (2003) 399–458

[13] Y. Fu, G. Guo, T. S. Huang, Age synthesis and estimation via faces: A survey, Pattern Analysis and Machine Intelligence, IEEE Transactions on 32 (11) (2010) 1955–1976.

[14] 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 8.1 (2017): 107-117.

[15] Jyoti S. Bedre ,Shubhangi Sapkal, : Comparative Study of Face Recognition Techniques: A Review. Emerging Trends in Computer Science and InformationTechnology–2012(ETCSIT2012) Proceedings published in International Journal of Computer Applications® (IJCA) 12.

[16] Li, Haoxiang, Zhe Lin, Xiaohui Shen, Jonathan Brandt, and Gang Hua. :A convolutional neural network cascade for face detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5325-5334. 2015.

[17] Zhu, Zhenyao, Ping Luo, Xiaogang Wang, and Xiaoou Tang. :Multi-view perceptron: a deep model for learning face identity and view representations. In *Advances in Neural Information Processing Systems*, pp. 217-225. 2014.

[18] Sun, Yi, Xiaogang Wang, and Xiaoou Tang. "Deep learning face representation from predicting 10,000 classes." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1891-1898. 2014.

[19] Jain, Vidit, and Erik Learned-Miller. Fddb: A benchmark for face detection in unconstrained settings. Vol. 88. Technical Report UM-CS-2010-009, University of Massachusetts, Amherst, 2010

[20] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097-1105. 2012.