ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 6, April 2017

Real-time Multiple Face Detection from Multiple Angles Using Multiple Haar Cascade classifiers and Convolutional Neural Network

Preeja Priji[1], Rashmi S Nair[2]

M.Tech Student, Department ofComputer Science, Mohandas College of Engineering, and Technology, Anad, Trivandrum, India [1]

Assistant Professor, Department ofComputer Science, Mohandas College of Engineering, and Technology, Anad Trivandrum, India[2]

***Abstract*— Haar Cascading is a type of object detection scheme used in image processing. Haar cascading uses haar features to find pattern and cascade the classifier to detect object from a scene. The haar cascade file contains the values of the haar feature for a specific shape. The training of haar cascade requires thousands of samples and time. Haar cascade file for frontal face and left profile faces are available to download. The haar cascading has less accuracy; hence it may include false positives. Other method like Deep learning doesnot produce false positives but are tedious process. In this work, we propose methods to detect multiple faces in different angles from a video using haar cascading and refine it with deep learning.**

***Index Terms*—Haar face detection, multiple faces, real-time face detection multiple angles, crowd analysis, uncontrolled face detection.**

I. Introduction

Image Processing is a very important field comprising thousands of useful and valuable applications such as face detection, recognition and processing is a fast growing filed in image processing, It is easy to determine a face if the individual is sitting at a normal frontal angle, but it is difficult to detect face from different angles. The haar cascading uses cascading of classifier trained by Adaboost method. The haar cascading method is faster compared to other neural network detection methods. Hence it can be used in real-time detection.

The goal in face detection is to identify and extract faces visible in an image[1]. Reliable face detection is one of the most studied research topics in the field of computer vision and precursor to face identification or matching. Haar Cascade classifier method is the commonly used method for face detection and detection of other objects. Haar Cascade XML contains the values for possible faces. Frontal xml contains values for frontal face and profile face file contains profile faces

values form left side.

The problem of haar cascading is the increased number of false positives in uncontrolled scenario. This can be rectified by further refining the detected faces by the Convolutional Neural
Networks (CNN).

In [2] the files from Face Detection Data Set and Benchmark (FDDB) dataset designed for studying the problem of unconstrained face detectionwas used and got very high accuracy whereaswith Open-I database, the result contained many false positives since Open I has uncontrolled scenario of faces.

The method [2] suggested is to use deep learning technique to the detected faces, so that the overhead of deep learning to scan entire image is reduced and scanning confined to the region that contains faces or false positive which are detected from Haar classifier. This paper is organized into three chapters, chapter II deals with extraction, detection and refining of faces, chapter III deals with result and analysis and chapter IVis the conclusion.

II. RELATED WORKS

The paper [3] present two methods for tracking pedestrians in videos with low and high density of crowds. For videos with low density crowd, [3] first individuals in each video frame is detected using a part-based human detector where occlusion is handled. In the second method, a global data association method based on Generalized Minimum Clique Graphs is used for tracking each.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 6, April 2017

Paper [3] present two approaches for tracking people in high density of crowds. In First method, scene layout constraint is captured by learning Dynamic Floor Field, Static Floor Field and Boundary Floor Field along with flow of crowd and it is used to track individuals in the crowd. In second method, the tracking is performed utilizing contextual and salient information.

This works [4]. focus on detecting and segmenting out crowds of humans from still photos. The goal is to determine if there is a crowd in a sample photo and if so, which portions of the image it includes the crowd. The detection of a crowd form a uncontrolled image environment is useful task in itself.crowd formation can cause delay in underground passages, shopping centers and pedestrian paths, or can an indication of civil unrest. The automotive industry, considers crowds of interest as a potential road hazard. Moreover, crowd segmentation is a useful pre-processing procedure that precedes higher level tasks, such as counting the number of individuals in crowd,and analyzing their behavioral dynamics and interaction. The application where crowd detection can be applied ranges from psychological research and macro-engineering, through to crime prevention and detection.
A crowd can be defined as a group of spatially proximate objects of a certain class. The work specifically considers human crowds, as the type that is usually of most concentrated in practice.

There are many reasons which makes crowd detection challenging. First, limited resolution of images that decreases the possibility of detection. Partial occlusions are prevalent in crowds, and the variation in dress, pose, light makes it difficult.Detection of individuals as the basic building element is not a promising approach [4]. Where asa method that directly looks for multiple people, faces problems of modelling an increased range of variability in their combinedappearance, also crowd specific factors such as the spacing of individuals in the crowd that is its density.

The work follows the above-mentioned line of work and extends it to the detection and tracking of people in high-density crowds instead of modelling individual interactions of people, the workuses information at a more global level provided by the geometry of scene and crowd density. Some crowd detection method avoids the hard detection task and attempt to infer person counts directly from low-level image measurements.These methods provide person counts in image regions but is uncertain about the location of detected faces.

The goaland contribution of the analyzed work is to combine these two sources of complementary information for improved person detection and tracking. The prediction behind the method is illustrated in Fig 3 where the constraints of person counts in local image regions helps improvement of the standard head detector.

The method is formulated in an energy minimization framework which combines crowd density estimates with the strength of individual face detections. This energy is minimized by jointly optimizing the density and the location of individual faces in the crowd. The work demonstrates optimization of such leads to significant improvements of state-of-the-art person detection in crowded scenarios with varying densities.

With crowd density cues, the constraints provided by scene geometry and temporal continuity of person tracks in the video is explored and demonstrate further improvements for person tracking in highly crowded scenario. The approach is validated on challenging crowded scenes from multiple video datasets.

The work [6] focuson developing effective features and robust classifiers for unconstrained face detection with arbitrary facial variations. Firstly, a simple pixel-level feature, called the Normalized Pixel Difference (NPD) is proposed. An NPD is the ratio of the difference between any two intensity values of pixel to the sum of the values. The NPD has several desirable properties, such as scale invariance, boundedness, and ability to reproduce the source image.It is easy to compute, involving only one addition, one subtraction, and one division between two valuespixels per feature computation.

Secondly a method to construct a single cascade classifier that can effectively deal with complex face contour and handle different pose and occlusions. The weak discriminative ability of NDP is solved by indicating that a subset of NPD features can be optimally selected by AdaBoost learning and combined to create discriminative features in regression tree which is a “divide and conquer” strategy to face and optimizedunconstrained face detection in a single classifier, without labelling the views in the training set of the face images. The proposed face detector is robust to pose variation, occlusion problem, and angular illumination, also to blur and low resolution image.

This work [7] mainly relies on a head detector to count people from a source image. For detecting the heads from the source image first the point of interest is detected using gradient information from the grey scale image. This

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 6, April 2017

approximately locates top portion of the head region to minimize the search region. The points of interest on the source image are masked using a foreground regional space obtained using background subtraction techniques including Vibesand Idiap. Then a sub-window is placed covering the points of interest based on information on perspective calibration and classifies as head or non-head region making use of a classifier. Multiple nearby detections are finally mergedto obtain result which is the no of faces.

III. EXTRACTING, DETECTION AND REFINING OF FACES.

The face extraction from an image is an easy process. We make use of 3 pass method to detect faces from front, left and right. The detection of faces from different angles from video is explained in 3 steps. Extraction, detection of face using Haar cascading and removal of false positives using Convolution Neural Network.

A. Extraction of image from video and preparation

First the first frame from a video file is extracted using Query Image function. The image file is first converted to grayscale, now the image is subjected to histogram equalization. The output image is ready to be analysed and is passed for detection. After detection, the next frame is extracted and the whole process is done for all frames.Fig 1 shows block diagram of image extraction and Fig 2 shows the image extraction stages.

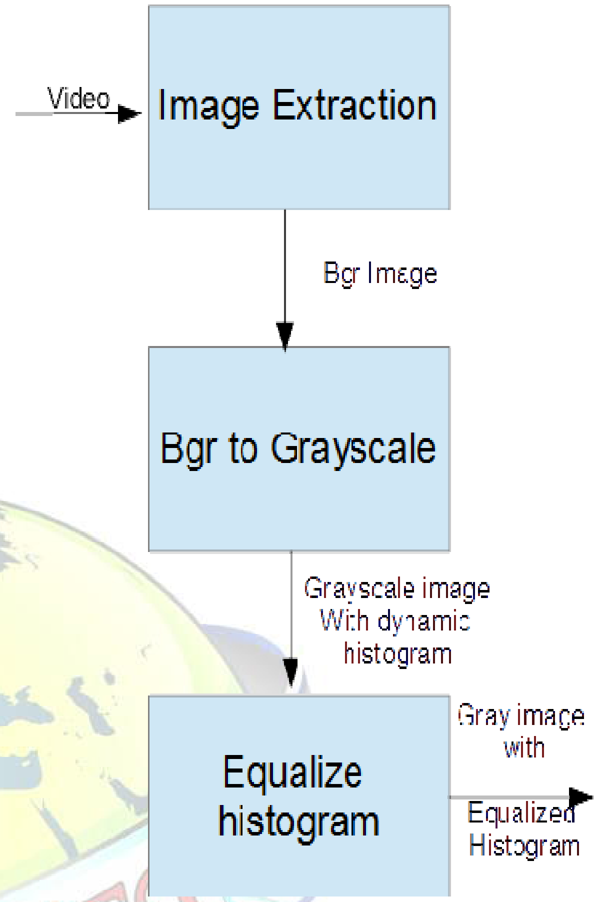

Fig. 1. Image extraction and preparation block diagram

Fig 2. (a) Extracted Image

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 6, April 2017

Fig 2. (b) Gray Scale Image,

Fig 2. (c) Histogram Equalized Image

B. Detection of faces using haar cascade classifier.

The detection process is a 3-stage process and is explained below. Fig 3 shows front, left and right profile faces.

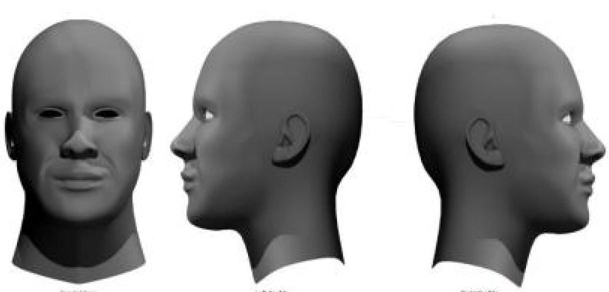

Fig 3. Front, Left and Right Profile faces

First, we detect all frontal faces using frontal face values. The detected face values are stored as rectangles in a list.

Fig 4 shows the detected frontal faces shown by red rectangle

Fig 4. Detected frontal faces shown by red rectangle

Now the same image is subjected profile face values. The left profile faces are detected in this step, the rectangles are compared with front faces list and checked for intersection. If intersection area is >0, the face is discarded since it is the duplicate of the frontal face already detected. The remaining values are appended to output face list. Fig 5 shows left profile faces detected in green rectangle.

Fig 5. Detected Left faces.

Now the image is flipped horizontally, and subjected to values in profile face again. Here right profile faces are detected, the detected rectangle's x coordinates are inverted using xinvert= width_of_rect - (x-1) [8].

The rectangles are checked for intersection with faces list and non-intersecting faces are added to list. Now output list

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 6, April 2017

contains all faces and false positives from the original image. This list is refined using CNN.

Fig 6 shows detected right faces in blue rectangle

Fig 6. Detected right faces from flipped image

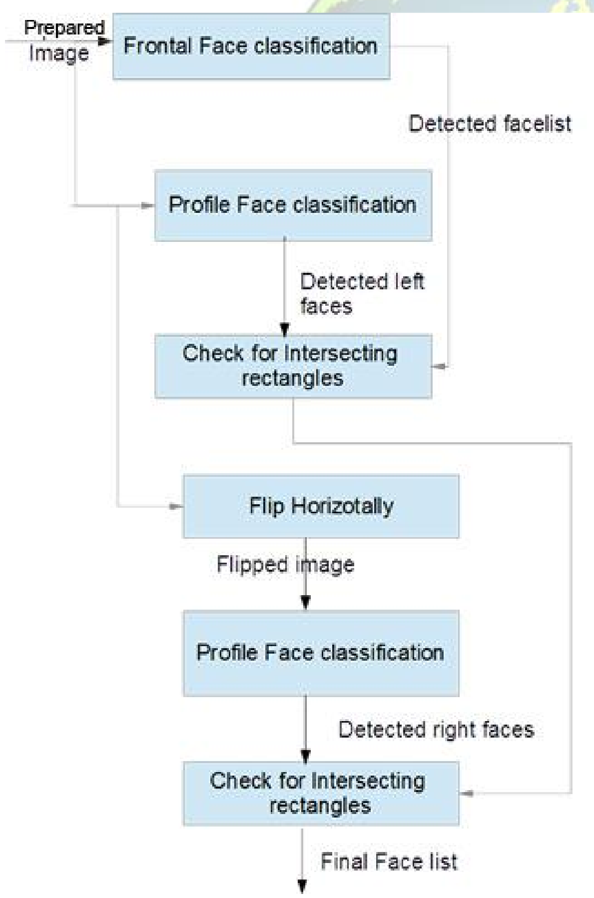

Fig. 7 Haar face recognition block diagram

C. Removal of false positives using CNN

The convolutional neural networks (CNN) are a special kind of multi-layer neural network designed for image processing. CNN explores spatial relationships of pixels in images to decrease of parameters in the neural network that must be trained [9][10].

There are four major idea used by CNNs:

1) Local connections: each unit in one layer is connected to a spatially-connected local subset of units in the previous layer;

2) Shared weights: all units in each of the feature maps in one layer have the same set of weights;

3) Pooling: a spatial sub-sampling step is applied to reduce the dimensions of the feature maps;

4) Many layers: the network may have more than 10 layers.

A CNN architecture consists of a number of convolutional and sub-sampling layers and severallayers that are fully connected [11]. The convolutional layer has several feature maps. Each unit connected to a local subset of units in the previous layer. That is each feature map is obtained by convolving the input with a linear filter then adding a bias and then passing through a non-linear function. The units in each feature map in the convolutional layer are calculated by sub-sampling layer[12] [13]. The process reduces the computational complexity for subsequent layers and provides a certain degree of shift-invariance. Multilayer perceptron (MLP) are fully connected network. The parameters of CNN are trained through back propagation algorithm. [5] proposed a principle in which the division is the urgent stage in iris acknowledgment. We have utilized the worldwide limit an incentive for division. In the above calculation we have not considered the eyelid and eyelashes relics, which corrupt the execution of iris acknowledgment framework. The framework gives sufficient execution likewise the outcomes are attractive.

Open source application cuda-convnet which uses NVIDIA GPUs to accelerate the computation speed can be used. [14] In cuda-convnet, the schemes of local normalization and overlapping pooling are used in a layer to improve generalization. A regularization method called dropout, whose key idea is to randomly drop units from the neural network during training is employed to reduce overfitting in the fully connected MLP layers. For the details of the architecture and the training protocol, refer to [15, 16].

Each of the faces detected are cropped from the source image and is fed to the CNN. CNN accepts if the given image values match the cumulative thresholds of then many to many neural connections the CNN rejects the face, the rectangle is removed from output list.

The Fig8 shows the input set of CNN.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 6, April 2017

Fig 8 Input of CNN

The block diagram shown below shows the working of CNN.

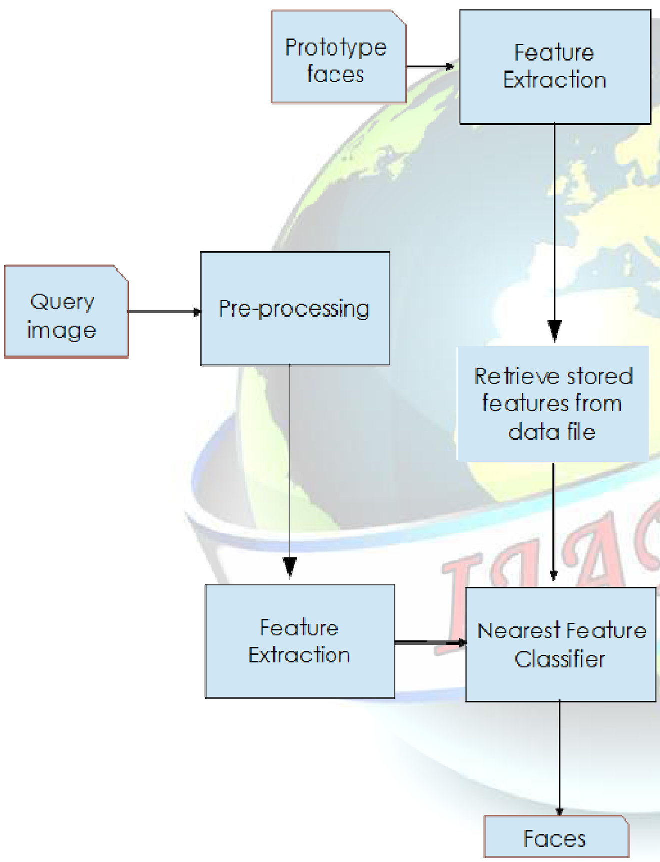

Fig. 9 CNN block diagram

The rectangles in output list are drawn to the source image and displayed. Fig 10. Shows the refined image

Fig 10. Image after removing false positives

IV. RESULT

The faces were detected and false positives were removed successfully. In picture containing 19 faces, 20 faces were detected, of these 3 were false positives. The CNN removed the three and 17 faces were successfully detected and marked [17] [18].

The accuracy of the method is estimated to be 94.53% on average in any scenario and number of false positives was 0. Speed of detection is significantly 10 times faster than CNN on whole picture [19]. Although detection of frontal face only is 3 times faster than detection of left right and frontal. The performance can be increased by using CUDA Libraries.Fig 11 shows Statistics graph of Haar detection Vs Refined technique and Fig 12 shows combined sample output of the process.

Below is the accuracy estimation table, accuracy statistics and screenshot of an image detected.

No of Faces in samples	Detected Faces	Number of False Positives	Accuracy
19	17	0	89.47
7	7	0	100
24	22	0	91.6
12	11	0	91.6
6	6	0	100

Table. 1. Accuracy statistics for refined method.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 6, April 2017

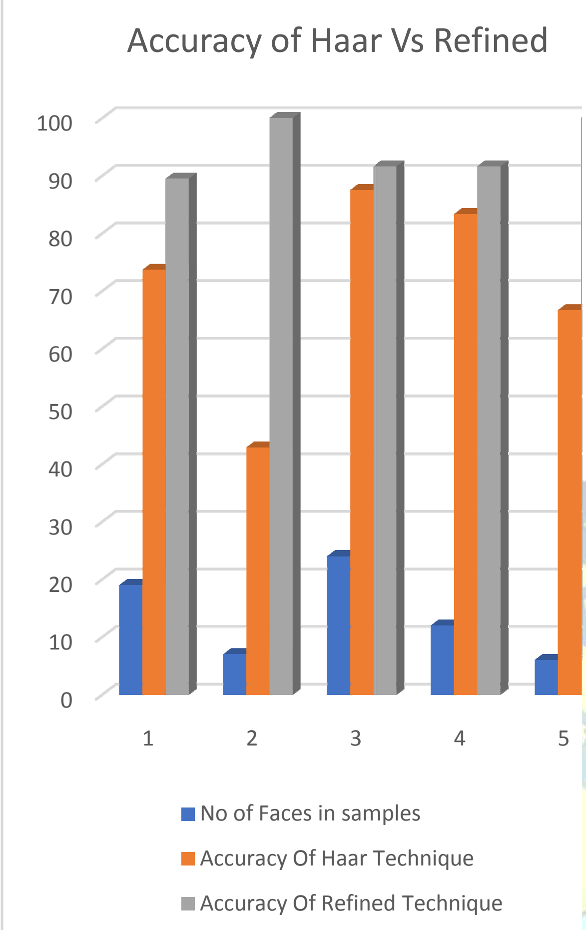

Fig 11. Statistics graph of Haar detection Vs Refined technique

Fig. 12 Sample output

V. CONCLUSION

The proposed method makes use of haar classification and CNN for an improved and more accurate faster face detection in real-time. The faces are detected from multiple angles with the use of haar cascade classifier. The method can be used in any complex scenario.

REFERENCES

[1] . Arel, I., Rose, D.C., Karnowski, T.P., “Deep machine learning – a new frontier in artificial intelligence research,” IEEE Computational Intelligence Magazine, 14 – 18 (November 2010).

[2] . Improving Face Image Extraction by Using Deep Learning Technique ZhiyunXue, Sameer Antani, L. Rodney Long, Dina Demner-Fushman, George R. Thoma National Library of Medicine, NIH, Bethesda, MD

[3] . AfshinDehghan, HaroonIdrees, Amir Roshan Zamir, and Mubarak Shah, “Automatic Detection and Tracking of Pedestrians in Videos with Various Crowd Densities”, Computer Vision Lab, University of Central Florida, Orlando, USA e-mail: adehghan@cs.ucf.edu

[4] . U. Weidmann et al. (eds.), Pedestrian and Evacuation Dynamics 2012, DOI 10.1007/978-3-319-02447-9__1,

[5] . Christo Ananth,"Iris Recognition Using Active Contours",International Journal of Advanced Research in Innovative Discoveries in Engineering and Applications[IJARIDEA],Volume 2,Issue 1,February 2017,pp:27-32.

[6] . Mikel Rodriguez1,4 Ivan Laptev2,4 Josef Sivic2,4 Jean-Yves Audibert3, Universite Paris-Est, Density-aware person detection and tracking in crowds, MPEG video compression standard, International Thompson publishing, 2010

[7] . Shengcai Liao, Anil K. Jain, Fellow, IEEE and Stan Z. Li, Fellow, IEEE, Unconstrained Face Detection 2013

[8] . LeCun, Y., Bengio, Y., Hinton, G., “Deep learning”, Nature, 521, 436-444 (2015).

[9] . Lienhart, R., Kuranov, A., Pisarevsky, V., “Empirical analysis of detection cascades of boosted classifiers for rapid object detection,” Proceedings of the 25th DAGM Symposium on Pattern Recognition. Magdeburg, Germany, (2003).

[10] . Garcia, C., Delakis, M., “Convolutional face finder: A neural network architecture for fast and robust face detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(11), 1408-1423 (2004).

[11] . Xue, Z., You, D., Chachra, S., Antani, S.K., Long, L.R., Demner-Fushman, D., and Thoma, G. R., “Extraction of endoscopic images for biomedical figure classification,” Proc. SPIE. 9418, Medical Imaging 2015: PACS and Imaging Informatics: Next Generation and Innovations, 94180P (March 2015)

[12] . LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., “Gradient-based learning applied to document recognition,” Proceedings of the IEEE, 86 (11), 2278–2324 (1998).

[13] . Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M.,

Berg, A.C., Fei-Fei, L., "ImageNet Large Scale Visual Recognition Challenge," International Journal of Computer Vision, (2015).

[14] . Zhang, C., Zhang, Z., "A survey of recent advances in face detection," Technical Report, MSR-TR-2010-66, Microsoft Research, (2010). [4] Viola, P., Jones, M.J., "Robust real-time face detection," International Journal of Computer Vision, 57(2), 137-154 (2004).

[15] . Krizhevsky, A., Sutskever, I., Hinton, G., "ImageNet classification with deep convolutional neural networks," Neural Information Processing Systems (NIPS), 1097-1105 (2012).

[16] . Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. "Dropout: a simple way to prevent neural networks from overfitting," The Journal of Machine Learning Research, 15(1), 1929-1958 (2014).

[17] . Rowley, H., Baluja, S., Kanade, T., "Neural network-based face detection", IEEE pattern Analysis and Machine Intelligence, 20, 22-38 (1998).

[18] . Milborrow, S., Morkel, J., Nicolls, F., "The MUCT landmarked face database," Pattern Recognition Association of South Africa, (2010).

[19] . Krizhevsky, A., Sutskever, I., Hinton, G., "ImageNet classification with deep convolutional neural networks," Neural Information Processing Systems (NIPS), 1097-1105 (2012).