Improvement of Face and Eye Detection Performance by Using Multi-task Cascaded Convolutional Networks

Mahmudul Hasan Robin[1], Md. Minhaz Ur Rahman[2], Abu Mohammad Taief[3] and Ms. Qamrun Nahar Eity[4]
Department of Computer Science and Engineering
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
mahmudulrobin17@gmail.com[1], minhaz.acps@gmail.com[2], taiefmaiden@gmail.com[3] and eity_cse@hotmail.com[4]

***Abstract*—Detection of face and eyes in unrestricted conditions has been a problem for years due to various expressions, illumination, and color fringing. Recent studies show that deep learning methods can attain impressive performance in the identification of different objects and patterns. As various systems may use the human face as input material, the increase in facial and eye detection performance has some significance. This paper introduces an enhanced face and eye detection technique through the use of cascaded multi-task convolutional networks for our dataset. We propose in this paper a deep cascaded multi-task system that exploits their inherent correlation to improve their performance. We collected 100 videos containing about 18265 images captured from our device and applied this dataset to the process and other systems proposed. The educated model was checked on our dataset and contrasted with the Haar cascade model as well. Our proposed method achieves a 98% percent accuracy rate considering our dataset which is superior to the other techniques used to detect the face and eye from an image. Besides, this paper also reflects a study of different methods of detecting the eye and face in tabular format from videos. The experimental results however indicate that the proposed approach demonstrates enhanced eye and face detection output from videos.**

***Index Terms*—MTCNN, Eye Detection, Face Detection, Haar Cascade**

I. Introduction

Face and eye detection have become important research topics in computer vision and pattern recognition [1], [2] in recent years because positions of the human face and eyes are essential information for many applications including psychological analysis [3], eye recognition, face recognition and facial expression analysis and medical diagnostics. Nevertheless, in many practical applications facial and eye recognition is quite difficult. The great visual variations of faces, such as occlusions, large pose changes, and intense lighting poses great challenges for these practices in real-world applications. The cameras are prone to light fluctuations and the distance from shooting, which makes the human eyes in a facial picture very excentric [4]. Sometimes the face is partly occluded, so we can not get a full facial image. For example, a cover test for detecting squint eyes covered half the face [5]. In this case, some existing methods of eye detection do not work, because they rely on facial model detection to spot the eyes. Although many methods have been developed to detect faces from photographs and eyes from facial images, one method that performs well in terms of accuracy, reliability, and efficiency is difficult to find. We are therefore trying to develop an effective and enhanced eye detection algorithm for our dataset to successfully detect eyes and faces with reasonable accuracy from the images.

In this paper, we suggested an effective and professional approach to help identify the face and eyes of videos without any human assistance based on both Haar Cascade and Cascaded Convolutional Networks using our dataset. In the future, our dataset will allow us to research on the behavioural posture and acts of human beings in diverse conditions. This paper indicates that the MTCNN method reliably outclasses other current approaches, using our dataset as a test dataset. In addition, counting the number of individuals on the scene by facial recognition of the face and eyes is also possible in a better way for certain purposes using the MTCNN method utilizing our dataset. The remainder of this paper is structured as follows: A summary of related works is presented in the Literature Review. We suggested Haar Cascade and MTCNN in the Proposed Methodology for our facial and eye detection dataset. Finally, findings are represented and discussed in the Experimental Results section for further review and decision making.

II. Literature Review

One of the most difficult and demanding tasks is to improve the accuracy of object detection in the computer vision field, such as the human face and eyes. Researchers around the world are working in this area to be able to use the best-found objects in several applications.

According to Kasinski [6] Haar cascade Classifiers are becoming common in face-end eye detection. It characterizes an HCC-based 3-stage hierarchical face and eye detection system. HCC is composed of 2500 positive facial expressions for identification of the face. There are 3500 images taken

978-1-7281-7366-5/20/$31.00 ©2020 IEEE

where there is no name. Face detectors are equipped with an image of 2500 left or right eyes and the images of the eyestrain negative sets. Total positive 94 percent and false-positive 13 percent are detected in facial detection. Eyes are detected at a rate of 88 percent with only 1 percent false positive outcome.

Based on deep convolutional networks approaches, Zhang [7] adopted three stages of deep convolutional networks that can predict the coarse-to-fine position of face and landmarks in a great manner. A recent study has shown that in this field, deep learning approaches can get enormous results. The author has suggested CNNs for eye detection consisting trio stages: Proposal Network (P-Net), Refinement Network (R-Net) and Output Network (O-Net). Experimental results finds these methods to exceed state-of-the-art methods over multiple demanding tests while preserving efficiency in real-time.

Lang Ye [8] suggested a novel CNN framework to boost the precision of eye detection directly utilizing the raw color values of image pixels by cascading two layers of CNN. The first point, senses rough bounding boxes of potential eye patches. The second step decides whether or not the rough bounding boxes belong to the eyes and exclude the non-eye bounding boxes. 8300 eye samples of different light circumstances, resolutions were obtained. Subsequently, entire samples were split into training and validation datasets of 500 samples per class in the validation set. The second stage of CNN outperforms the first tier of CNN, reaching an accuracy rate of 73 percent and a recall rate of 76 percent respectively.

III. Proposed Methodology

There are two distinct models in our proposed approach for defining the face and eyes of our dataset. First model based on the MTCNN approach to detect human objects while the second model is Haar Cascade Algorithm, which also detected eyes and face from our dataset. The use of MTCNN gives our dataset a better result.

A. Proposed Methodology of Face and Eye Detection using MTCNN

The eyes and face detection had been achieved using multi-task cascaded convolutional networks in our first prospective model. Our proposed face and eye detection system consists of three key stages: the selection of data sets, preprocessing, face and eye detection using MTCNN. The outcomes of our study worked satisfactorily. The key stages of our proposed model "Fig. 1" are outlined below.

1) Dataset Collection: Our dataset was created, and the dataset consists of 100 videos. Those videos were taken from volunteers. Light conditions were true and all the videos were taken from a single phone that prioritized voluntary frontal facial vision. We also included people having abnormal eye position. Each video contains around 150-200 frames and is 6-10 seconds long. Besides, each video was taken using a 12-megapixel cell phone with a resolution of 720x 1080 pixels. The distance from the camera to the person is around 1 to 3 meters. Increasing the distance will make the system unreliable, and so we've taken all the videos from close up.

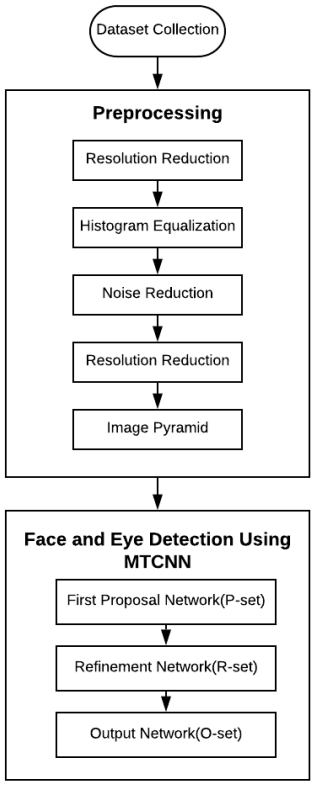

Fig. 1. Flowchart Diagram for Proposed Method.

2) Preprocessing: Preprocessing is a key part of image processing. Every frame of each video has a resolution of 720x 1080p.The histogram was drawn to find the lightness distribution of each container. Instead, noise reduction is carried out with the use of other morphological operations: closing, erosion, and opening. We created an image pyramid, which is the input for detecting eyes using MTCNN.Then, we re dimension the pyramid to various scales.

3) Face and Eye Detection Using MTCNN: CNN or Convolutional Neural Network is an evolved and feed-forward artificial neural deep learning network for visual image analysis. After the images are pre-processed, the input images are used with MTCNN to identify the eyes and face. There are three steps [7] it has to go through:
   a) Stage-I: It is called the Proposal Network (P-Net) that obtains the windows of the candidate and the vectors of their boundary box. Then we use the bounding box regression vectors which are estimated to tweak the members. Non-maximum suppression (NMS) will then be used to combine

heavily mirrored members.

b) Stage-II: All members are transmitted to the next CNN, named Refine Network (R-Net), which also eliminates a significant number of false candidates, conducts boundary box regression verification, then merges members for NMS.

c) Stage-III: This stage's much like the previous stage but it appears to explain the face in more detail at this stage. A circle is drawn particularly to highlight and finally detect the eyes from the face.

B. Face and Eye Detection Applying Haar Cascade Algorithm

Haar-based Cascade Classifiers is also an effective technique of identification objects. There are several ways to detect face and eye from images using har cascade classifier. For example, face detection using three additional weak classifiers which are then applied to basic Hair-like features focused on cascaded classifiers [9]. However, by applying the Haar Cascade algorithm to our dataset, the program can first detect the name. Second, after the face has been identified, the eye would be identified in the face area [10]. These apps evaluate the discrepancy in contrast values among neighboring rectangular pixels sets, instead of using a pixel's intensity scores. Double or triple neighboring clusters with a relative variance of contrast form a Haar-like feature. Uses haar-like features to detect an object. Haar traits can be measured by raising or lowering the scale of the pixel cluster being studied. It allows use of software to monitor objects of various sizes. Face and eye cascade of haar-cascade classifier was used to identify the facial and eye from numerous frames.

IV. Experimental Results

Our observation is performed in a pair of different model categories. The first experiment is carried out using the MTCNN model, where we used our dataset. The second experiment is also performed using the Haar Cascade Algorithm on our dataset. To explain our proposed model, a comparative review of our proposed face-and-eye recognition models using MTCNN and Haar Cascade is shown after our dataset is implemented as a test data collection. Use MTCNN we got 98% percent accuracy, and using Haar Cascade 68.16% accuracy is achieved. Then we compared our results from both of our experiments according to precision, accuracy, and recall. Finally, we equate our work in the relevant field of object detection with the existing works. Section A to C provides a comprehensive overview of our experimental results and a summary of our findings.

A. Results of MTCNN & Haar Cascade Algorithm

After applying our dataset to the MTCNN process, we determined the eyes and face of the images for about 100 videos at a rate of 98%. Here, the result indicates that a significant outcome has been achieved using Multi-task Cascaded Convolutional Networks. "Fig. 2", displays some of the images from our dataset where the eyes and face were successfully determined using the MTCNN process.

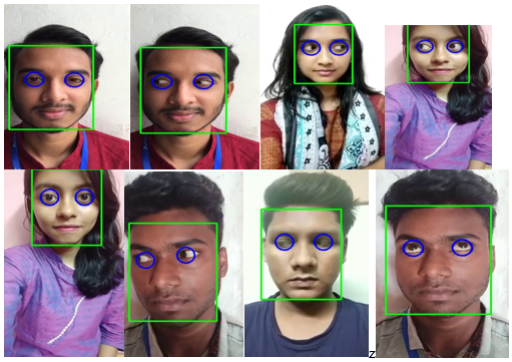

Fig. 2. Face and Eyes Detection Using MTCNN

After applying our dataset to the Haar Cascade method, we determined the eyes and face at a rate of 68.16%. "Fig. 3", displays some of the images from our dataset where the eyes and face were successfully determined using the Haar Cascade method.

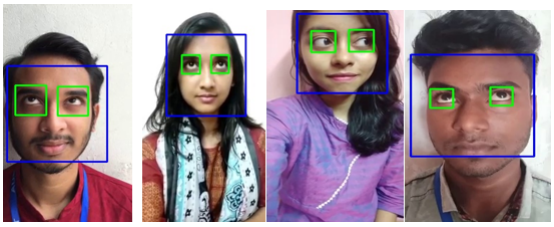

Fig. 3. Face and Eye Detection Using Haar Cascade

B. Result Comparison

We can analyze the detection rates among these methods using Different Datasets in different methods. We have evaluated the difference in detection rate between the methods in "Table. I".

TABLE I
COMPARISON OF DETECTION RATE OF EYES AMONG VARIOUS METHODS FOR DIFFERENT DATASETS

Algorithm	**Detection rate**	**Dataset**
Viola Jones	61.81%	MIT
Hough Transform	83%	Casia Database
MTCNN	92%	NICE-II and MICHE database
Haar Cascade	94%	set of 10000 test images

Using our dataset in different methods we can analyze the detection rates among these methods. In the "Table. II", we have analyzed the difference in detection rate among the methods.

TABLE II
COMPARISON OF DETECTION RATE OF EYES AMONG VARIOUS METHODS FOR OUR OWN DATASETS

Algorithm	Detection rate
MTCNN	98%
Haar Cascade	68.16%

Various datasets are used to compare between different methods. Hence, drawing a comparison chart "Fig. 4" can compare the detection rate between our own dataset and different datasets for different methods.

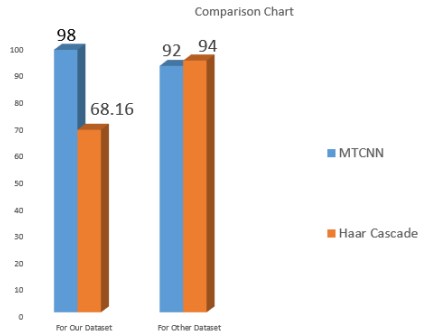

Fig. 4. Comparison Amongst Different Methods for Our Own Dataset and Other Different Datasets

C. Analysis & Discussion

We have calculated the confusion matrix of eyes and face detection for our proposed method MTCNN using our dataset.

TABLE III
CONFUSION MATRIX OF FACE AND EYES DETECTION(MTCNN)

	Positive	Negative
Predicted Positive	17620	335
Predicted Negative	30	280

It is drawn in "Table. III". The confusion matrix has been generated from 100 videos containing about 18265 images.

TABLE IV
CALCULATION OF CONFUSION MATRIX FOR FACE AND EYES DETECTION(MTCNN)

Measure	**Value(%)**	**Derivation**
Precision	98.13	PPV = TP / (TP + FP)
Recall	99.83	TPR = TP / (TP + FN)
Specificity	45.53	FPR = FP / (FP + TN)
F-Measure	98.97	F1 = 2TP / (2TP + FP + FN)
Accuracy	98.00	TP+TN / TP+TN +FP+FN

And according to "Table. IV" the accuracy rate is 98%. Where the precision is 98.13% and recall is 99.83%. and finally, the F-1 Score is 98.97%.

After evaluating all the comparisons, it is clear that our data set is better for deep learning methods(MTCNN) to detect eyes and face. Therefore, given all the different methods for our own dataset and other datasets, the overall performance in the deep learning process was satisfactory.

V. CONCLUSION & FUTURE WORK

In this paper, we eventually proposed a framework for face and eye detection based on multi-task cascaded CNNs. Experimental results and our other proposed framework show that our MTCNN methods consistently outperform most of the existing methods, considering our dataset as a test dataset. There are plenty of applications that can use our proposed technique like our dataset since we can count the number of people in a scene by image recognition of the face and eyes. In the future, we will try to work on other eyes and face research activities such as facial expression or attitude identification, detection of fatigue, movement of the iris and detection of a questionable observer. We have a plan to work on these tasks by using our dataset so we can use the data set properly to train our model for performance improvement.

REFERENCES

[1] H. Fu, Y. Wei, F. Camastra, P. Arico, and H. Sheng, "Advances in Eye Tracking Technology: Theory, Algorithms, and Applications," Computational Intelligence and Neuroscience, vol. 2016, Article ID 7831469, 2016.

[2] L. Zhang, Y. Cao, F. Yang, andQ. Zhao, "Machine Learning and Visual Computing," Applied Computational Intelligence and Soft Computing, vol. 2017, Article ID 7571043, 2017.

[3] D. Schneider, A. P. Bayliss, S. I. Becker, and P. E. Dux, "Eye movements reveal sustained implicit processing of others mental states.," Journal of Experimental Psychology: General, vol. 141, no. 3, pp. 433–438, 2012.

[4] B. Li and H. Fu, "Real Time Eye Detector with Cascaded Convolutional Neural Networks," Applied Computational Intelligence and Soft Computing, vol. 2018, pp. 1–8, 2018.

[5] L. Birgit, "Pediatric Ophthalmology, Neuro-ophthalmology, Genetics: Strabismus-new Concepts in Pathophysiology, Diagnosis, and Treatment," in Pediatric Ophthalmology, Neuroophthalmology, Genetics: Strabismus-new Concepts in Pathophysiology, Diagnosis, and Treatment, M. C. Brodsky, Ed., Springer Science Business Media, and Treatment, 2010.

[6] A. Kasinski and A. Schmidt, "The Architecture of the Face and Eyes Detection System Based on Cascade Classifiers," Advances in Soft Computing Computer Recognition Systems 2, pp. 124–131, 2007.

[7] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499–1503, 2016.

[8] L. Ye, M. Zhu, S. Xia, and H. Pan, "Cascaded Convolutional Neural Network for Eye Detection Under Complex Scenarios," Biometric Recognition Lecture Notes in Computer Science, pp. 473–480, 2014.

[9] L. Cuimei, Q. Zhiliang, J. Nan, and W. Jianhua, "Human face detection algorithm via Haar cascade classifier combined with three additional classifiers," 2017 13th IEEE International Conference on Electronic Measurement Instruments (ICEMI), 2017.

[10] N. L. Fitriyani, C.-K. Yang, and M. Syafrudin, "Real-time eye state detection system using haar cascade classifier and circular hough transform," 2016 IEEE 5th Global Conference on Consumer Electronics, 2016.