Eye Movement Classification Using CNN

Milu Prince
Dept of Computer Science And Engineering
Rajagiri School Of Engineering And Technology
Kochi, India
miluprince98@gmail.com

Neha Santhosh
Dept of Computer Science And Engineering
Rajagiri School Of Engineering And Technology
Kochi, India
nehasanthosh98@gmail.com

Nimitha Thankachan
Dept of Computer Science And Engineering
Rajagiri School Of Engineering And Technology
Kochi, India
nimithathankachan21@gmail.com

Reshma Sudarsan
Dept of Computer Science And Engineering
Rajagiri School Of Engineering And Technology
Kochi, India
lechusudarshan@gmail.com

Ms.Anjusree V.K
Asst. Professor, Dept of Computer Science And Engineering
Rajagiri School of Engineering And Technology
Kochi,India
anjusreevk@rajagiritech.edu.in

***Abstract*—In today's world eye has significant importance in technological advancements. Eye contact and eye gaze are an important factor for communication. Eye movement has a wide range of applications in various technology. In recent years, eye movement detection is gaining a lot of attention. Our aim is to create an approach that employs deep convolutional neural networks which will run for both eyes in parallel for visual feature extractions. Dataset images for training and validation were obtained from the eye chimera dataset. Testing is done by inputting real-time videos.**

***Index Terms*—face detection, eye detection, eye movement classification**

I. Introduction

Eye-tracking is a process where the eye gaze is measured to understand where a person is looking and also to analyze various cognitive processes of a person. There are a number of methods that are commonly used for eye movement detection and classification. Eye movements can be broadly classified into left, right, up, down, and center. It has wide importance in different industries, medical field [18], gaming, etc. For example we can evaluate if a person is telling truth or not by the movement of the eyes [2], also the diagnosis of various diseases such as Alzheimer's can be done. Eye-tracking is widely used in human-computer interaction [11].Eye-tracking technology can be a boon in the life of physically challenged people [14]. Deep convolutional neural networks are being used for image classification and detection problems.

The aim of this project is to employ deep neural networks that can detect various types of eye movements in a video. The different types of eye movements to be classified are up, down, right, left, center and blink eye movements. Here we employ deep neural networks that run for both eyes in parallel for visual feature extraction using a preprocessed dataset of different eye movements in a video.

There are various design and implementation constraints that include unavailability of predefined data set, high quality camera is required, training data model needs to be updated every 24hrs in colab. The assumptions and dependencies are camera will capture clear images from which we can extract eye features, then on using GPU in colab more processing power is there compared to CPU, sufficient lighting conditions are available, also sufficient dataset is available for training.

II. Literature Survey

Eye movement and classification can be done using various methods. For eye detection and classification as an initial step it is required to detect the face which can be done by various methods like Haar cascade, HoG and Neural network. Some of the methods like Haar cascade, Facial landmarks and Harris Corner Detector, Sobel Edge Operator and many other methods are used for eye detection purposes. The different classification methods used are CNN, SVM, Electrooculography and so on.

A. Face Detection Methods

Face detection is a technology that helps you to identify the human face in a digital image. The process of face detection is complicated due to the variations present across the human face. Haar cascade based face detection is a machine learning approach which is used to train the classifier and extract features from it. Another face detection method is HOG face detector in dlib. Histogram of oriented graphics(HOG) is a feature descriptor used for the purpose of object detection. It counts the occurrences of gradient orientation in a localized portion of the images. HOG face detector will not detect face at odd angles. It is a good frontal face detector. Then neural

978-1-7281-6453-3/20/$31.00 ©2020 IEEE

networks are also used to detect the face. Neural network examines small windows of an image, and chooses whether each window contains a face or not.

B. Eye Detection Methods

Haar cascade is one of the methods used for eye detection. The algorithm requires images with and without faces that is positive images and negative images to train the classifier and then to extract features from it [19]. Each feature selected seems to focus on the property that the region of the eyes is often darker than the region of the face. Hough transform is another method for eye detection. In this method it automatically detects the face in image and marks with a circular pattern using the Hough Transform. Hough transform is an algorithm that detects edges, more specifically used in detection of lines and circles in many kinds of images. Detection is possible even under noise. The circles are used to detect the eyes and sometimes it may detect unwanted locations also, so a series of tests are done to check conditions and remove the unwanted detected circles. The first test is the Horizontal test, then symmetric test, upper half test and dark pixel test is done in sequence [19]. Facial landmarks is another technique were facial landmarks are detected using the dlib library which uses a pre-trained face detector which is based on a modification to the Histogram of Oriented Gradients and Linear SVM method for object detection. There are 68 landmarks of the face. The landmark coordinates from 36 to 41 is the first eye and 42 to 47 is the second eye. For every point there is a specific index assigned. Left eye points: (36, 37, 38, 39, 40, 41) and Right eye points: (42, 43, 44, 45, 46, 47). Harris corner detector and sobel edge detector can also be used for eye detection. Here we convert the input human face image in RGB colour into grayscale image to perform various operation for proper eye detection. Harris corner detector is been used to detect the corners on the images of eyes. The corner detection block finds the corners in the image based on the pixels that have the largest corner metric values. The operator like sobel edge has been used to detect the edges on obtained extracted eye region image.

C. Eye Classification Methods

One of the popular eye classification method used is CNN. Convolutional neural network is a type of artificial neural network. It is a type of feed forward neural network [7]. Many computer vision tasks use these networks even though they require large training time since they provide high accuracy rate. Convolutional neural networks can be used for the classification of images [5]. In eye movement classification also convolutional networks can be used. Two convolutional neural networks can be used parallel for both left and right eyes separately [1]. Electrooculography is another method for eye classification. Electrooculography measures the electric potential between the front and back of the eye. An electric field exists near the cornea and retina. As the eye ball rotates there will be change in the electric field, and these electric fields are amplified and analysed to deduce the eye position.

III. Hardware and Software Requirements

A. Python 3.6 and Google Colab

Python 3.6 has been used as the programming language. It provides a wide range of libraries for doing the deep learning projects. Colab is a free cloud service based on Jupyter Notebooks for machine learning education and research. It provides a runtime fully configured for deep learning and free-of-charge access to a robust GPU.

B. OpenCV(Open Source Computer Vision Library)

OpenCV is a software library which is built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products [3].

C. Deep Convolutional Neural Network

Convolutional neural networks is a neural network which is used to classify images and to cluster images on their similarities. It is also used to perform detection of objects in an image. CNN is a Deep Learning algorithm which takes an input image and then based on weights of objects in the image we can differentiate one object from other that is to classify objects in a image [5]. It mainly consist of an input layer, an output layer and various hidden layers. Convolutional neural network can be used to identify patterns in images which is done by convoluting over an image and looking for patterns [10]. The lines and corners are found out in first few layers of the Convolutional Neural Network. These patterns are further passed down through our neural network which will then start recognizing more complex features as it go deeper. This makes CNN preferably good at identifying objects in images. CNN contains various layers such as convolutional layer, pooling layer, activation layer.

D. Histogram equalization

Histogram equalization increases the global contrast of the images. The intensities can be better distributed on the histogram [17]. This allows for areas of lower local contrast to gain a higher contrast. Histogram equalization accomplishes this by effectively spreading out the most frequent intensity values.

E. Camera

By using good quality camera , the eye detection can be done properly.

F. Dlib Library

Dlib is a modern toolkit containing machine learning algorithms and tools for creating complex software to solve real world problems. It provides complete and precise documentation for every class and function [9]. It allows the user to easily implement face detection, face recognition and even real-time face tracking. Dlib library is used for face detection and facial landmark detection. The frontal face detector function of dlib is used in order to detect the face. This detector is based on histogram of oriented gradients (HOG) and linear SVM.

IV. EXPERIMENT

The proposed method is to employ deep neural networks which can detect various types of eye movements in a video. The different types of eye movements to be classified are up, down, right, left, centre and blink eye movements [1].The system architecture is shown in the Fig. 1.

A. Face Detection and Eye Detection in Video

The face detection and eye detection is done using the dlib library and OpenCV. The frontal face detector function of dlib is used in order to detect the face. We also use the facial landmarks to detect eye from the face. The file for facial landmark detection is "shape predictor 68 face landmarks.dat" [15]. We process the video frame wise in order to detect the face and eyes. Convert the frame to grayscale image and use the function to detect the face. Use the facial landmark function in order to detect the eyes. The location of eyes in the facial landmark file is in points : Left eye(36,37,38,......41) and right eye(42,43,......47) which is shown in Fig. 2.

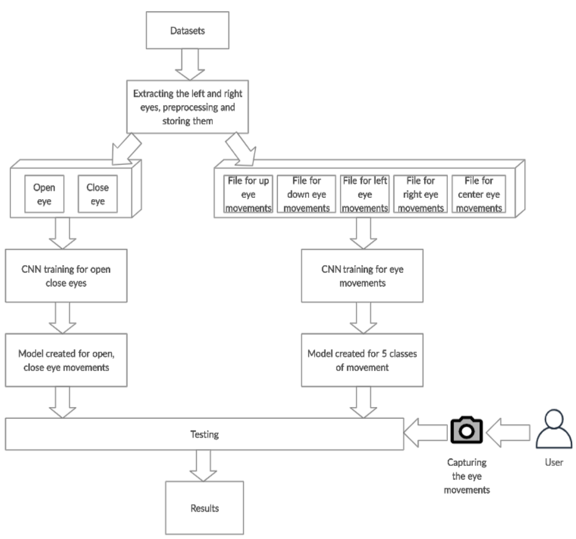

Fig. 1. System Architecture

Fig. 2. Facial Landmarks

B. Dataset Creation and Pre-processing

The Eye Chimera dataset which was available in the internet is used in this project. Eye chimera dataset has images of people's faces with different types of eye movements. After detecting the eyes from the face, both eyes were cropped separately from the images and were resized to size 50x50 pixels. Since the dataset did not have sufficient amount of images for each type of eye movement, we clubbed the cropped right and left eyes together. By doing this we could double the number of images for each eye movement class. Data augmentation techniques such as image flipping was done. When flipped, right eye's right movement will change to left eye's left movement and vice versa . Similarly , when flipped, the right eye's left movement will change to left eye's right movement and vice versa. This in turn helped in increasing the images in the dataset further. Gray scaling was done for faster training time. Histogram equalization was done to improve contrast in the images. This helped in better understanding of the eye position in the images. For blink detection model, we used the open eye image dataset created manually along with the open eye images from the eye chimera dataset. For closed images we used Closed Eyes in the Wild(CEW) dataset along with some images created by ourselves. Resizing and histogram equalization was done for these datasets also.

C. Training of the model

1) Blink detection Model: The dataset for open and closed eye images contain 1654 images separately. We splitted the preprocessed image set into 80 percentage for training and the remaining to validation and test set [8]. The input stage consists of the two classes of eye images open and close of size 50x50 pixels. Here the CNN consists of 3 convolution stages. Each stage will be followed by Rectified Linear Unit(ReLu) [7]. ReLu inntroduces a non-linearity to the activations. Then a max pooling layer is added after the ReLu stage. After convolution, ReLu and max pooling in the third convolution stage, the outputs are joined in fully connected layers. For the final layer a sigmoid activation function is used to classify to the two classes [6].The architecture of the model is shown in Fig. 3.

2) Left, Right and Centre classification Model: Here the input for the CNN is the 3 classes of images.That is, the images of left, right, centre movements of eyes [4].The dataset consists of 536 preprocessed images for the 3 classes separately. This model consists of 3 convolution stages. Each stage will be followed by Rectified Linear Unit(ReLu) [13]. ReLu layer introduces a non-linearity to the activations. Then a max pooling layer is going to be added after the ReLu stage. Then two similar covolution stages will be also added. After convolution, ReLu and max pooling in the third convolution stage, the outputs should be joined in fully connected layers. The final layer activation function is a softmax function [6]. The class with the highest prediction score will be selected as the predicted class [16].The architecture for the left, right, centre eye movement classification is shown in Fig. 4.

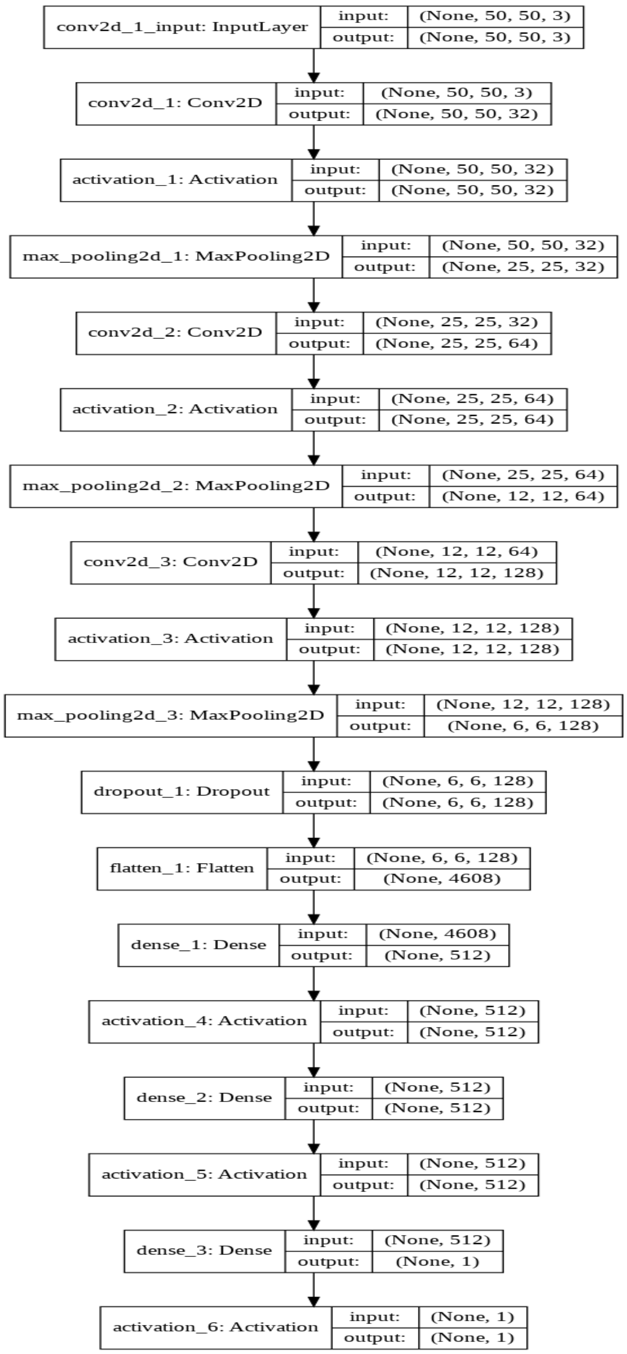

Fig. 3. Open and Close eyes classification model

3) Left, right , centre, up and down classification model: We added two more classes which is up and down movements [4]. For up and down classes since the dataset was not so accurate, we added some more images into the dataset manually. For creating this model, we used the same architecture of the model which was used for training the blink detection but with a softmax function at the final layer.The class with the highest

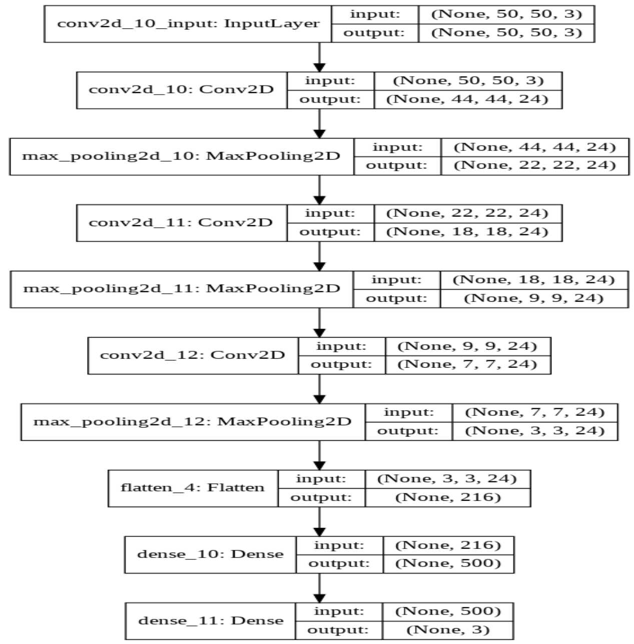

Fig. 4. Left,Right and centre eye movement classification model

prediction score will be selected as the predicted class.

D. Testing

Here we tested for blink detection with 3 class(left, right, centre) classification and also blink detection with 5 class(left, right, centre, up and down)classification.

Video is captured through webcam for testing the blink detection. Two counters which are memory counter and blink counter were initialized to 0. Here the video is splitted into frames and eyes are detected from the frames. The eyes are then preprocessed by resizing to 50x50 pixels and applying histogram equalization and then fed into the model. Left eye is fed into the trained model for predicting open and close. A predicted score is obtained.Right eye is fed into the trained model for predicting open and close. A predicted score is obtained. Calculate the combined average predicted score for the right and left eyes. If score is less than or equal to 0.5, image belongs to class CLOSE then increment the memory counter.Otherwise image belongs to Class OPEN. If memory counter is between a particular value, we increment the blink counter and set memory counter=0 [8]. If the eyes are predicted as open, the left and right cropped eyes are fed into eye movement classification model.The individual scores from left and right eyes for each class are combined by taking average.According to the scores the movement are classified as centre,up,down,right or left [4].

V. Results and Observation

The results obtained from the experiment are shown in table 1. The classification accuracies obtained for the open and close eye classification model,left, right, centre classification model and left, right, centre, up and down classification model are shown in the table 1. For the open and close eye classification, the CNN model was trained for 15 epochs and drop out of 0.5 was done in order to avoid over-fitting as the dataset was not huge.For the left, right, centre eye movement classification, the CNN model was trained for 50 epochs.For the left, right, centre, top and down eye movement classification, the CNN model was trained for 200 epochs.

The classification accuracy of open and close eye classification model which was used for blink detection is 98 percent. The validation accuracy for this model is 99 percent. The classification accuracy of the 3 class(left, centre, right) classification model was higher than that of the 5 class(left, centre, right,up,down) classification model. The accuracy in 3 class classification is 96 percent.The validation accuracy in this case is 98 percent. The accuracy

TABLE I
RESULTS TABLE

Classification type	Accuracy
Open and close eye	98%
Left, right, centre eye movements	96%
Left, right, centre, up and down eye movements	90%

in 5 class classification is 90 percent. The validation accuracy in this case is 89 percent. The prediction scores from both eyes are used to obtain the final output prediction. The classification without using spectacles was more accurate than using spectacles. The histogram equalization done in the images helped to increase the accuracy as it helps to differentiate the iris and the sclera more, which plays a major role in movement classification. The classification is also depended on the lighting conditions and the quality of the camera used in the experiment.

VI. Conclusion

Nowadays, the use of deep learning technology has increased to high level especially in the real world. Eye tracking and eye movement classification tasks have a special role in various fields especially medical and health related sectors which give an insight about a person's cognitive process. Image classification and further image detection is possible with the growth and development in neural networks. Deep neural network is used for eye movement classification that classifies movements as up, down, left, right movement or blink movements.

In brief here an input video captured is divided into frames, from which the face and eye detection is performed and after which the movements are classified. Face detection is done by using dlib frontal face detector and eye movements are classified using dlib facial landmarks.

VII. Scope for future work

The existing design can be modified and can be made better by incorporating RNN that is Recurrent Neural Network which helps in making time series predictions as RNN is dependent of time [12]. Also by increasing the number of datasets, the accuracy of the model can be further increased and by making use of the classified images we can implement and develop some real world applications in future.

References

[1] Marko Arsenovic, Srdjan Sladojevic, Darko Stefanovic, and Andras Anderla. Deep neural network ensemble architecture for eye movements classification. In *2018 17th International Symposium INFOTEH-JAHORINA (INFOTEH)*, pages 1–4. IEEE, 2018.

[2] Maria C Bradley, Donncha Hanna, Paul Wilson, Gareth Scott, Paul Quinn, and Kevin FW Dyer. Obsessive–compulsive symptoms and attentional bias: An eye-tracking methodology. *Journal of behavior therapy and experimental psychiatry*, 50:303–308, 2016.

[3] Gary Bradski and Adrian Kaehler. *Learning OpenCV: Computer vision with the OpenCV library*. " O'Reilly Media, Inc.", 2008.

[4] Anjith George and Aurobinda Routray. Real-time eye gaze direction classification using convolutional neural network. In *2016 International Conference on Signal Processing and Communications (SPCOM)*, pages 1–5. IEEE, 2016.

[5] Sabrina Hoppe and Andreas Bulling. End-to-end eye movement detection using convolutional neural networks. *arXiv preprint arXiv:1609.02452*, 2016.

[6] Nadia Jmour, Sehla Zayen, and Afef Abdelkrim. Convolutional neural networks for image classification. In *2018 International Conference on Advanced Systems and Electric Technologies (IC_ASET)*, pages 397–402. IEEE, 2018.

[7] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1725–1732, 2014.

[8] Ki Wan Kim, Hyung Gil Hong, Gi Pyo Nam, and Kang Ryoung Park. A study of deep cnn-based classification of open and closed eyes using a visible light camera sensor. *Sensors*, 17(7):1534, 2017.

[9] Kyle Krafka, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, and Antonio Torralba. Eye tracking for everyone. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2176–2184, 2016.

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

[11] Päivi Majaranta and Andreas Bulling. Eye tracking and eye-based human–computer interaction. In *Advances in physiological computing*, pages 39–65. Springer, 2014.

[12] E Yu Malakhova, E Yu Shelepin, and RO Malashin. Temporal data processing from webcam eye tracking using artificial neural networks. *Journal of Optical Technology*, 85(3):186–188, 2018.

[13] Chunning Meng and Xuepeng Zhao. Webcam-based eye movement analysis using cnn. *IEEE Access*, 5:19581–19587, 2017.

[14] Omkar Prabhune and Priti P Rege. Speakingeyes: Enabling paralyzed people to communicate. In *2019 IEEE 16th India Council International Conference (INDICON)*, pages 1–4. IEEE, 2019.

[15] S Sharma, Karthikeyan Shanmugasundaram, and Sathees Kumar Ramasamy. Farec—cnn based efficient face recognition technique using dlib. In *2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)*, pages 192–195. IEEE, 2016.

[16] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[17] JM Soha and AA Schwartz. Multispectral histogram normalization contrast enhancement. 1979.

[18] Susan Stuart, Lisa Alcock, Alan Godfrey, Stephen Lord, Lynn Rochester, and Brook Galna. Accuracy and re-test reliability of mobile eye-tracking in parkinson's disease and older adults. *Medical engineering & physics*, 38(3):308–315, 2016.

[19] Phillip Ian Wilson and John Fernandez. Facial feature detection using haar classifiers. *Journal of Computing Sciences in Colleges*, 21(4):127–133, 2006.