International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 10 | Oct 2022 www.irjet.net p-ISSN: 2395-0072

An Eye State Recognition System using Transfer Learning: Inception‑Based Deep Convolutional Neural Network

Jigar Sapkale[1], Dr. Surendra J Bhosale[2], Dr. Rahul Ingle[3]

[1]M. tech Student, Dept. of Electrical Engineering, VJTI, Maharashtra, India
[2] HoD, Dept. of Electrical Engineering, VJTI, Maharashtra, India
[3] HoD, Diploma, VJTI, Maharashtra, India

Abstract - *This study proposes a deep convolutional neural network (DCNN)-based technique for eye state identification (closed or open) utilizing the MRL datasets. goal, pre-trained CNN architectures on INCEPTION were first trained on datasets, which included open and closed-eye states, before being evaluated and their performance quantified. For this study we are using inception v3 architecture. Simultaneously, the DCNN architecture suggested on the MRL datasets has been shown to be an appropriate and effective approach for eye state recognition based on the results compared to earlier research. Through expanding to the literature on eye state identification, this approach may help with the creation of HMI systems.*

Key Words: **Human–machine interaction, Deep learning, Deep convolutional neural network, Transfer learning, Inception**

1. INTRODUCTION

The condition of the eye is among the facial characteristics used to determine whether an eye is open or closed. It is also a critical requirement for correctly depicting a person's physiological condition. Despite the fact that eye condition can be expressed in a wide range of ways, it can be broadly split into two categories: open and closed. It has a lot of promise in fields like sleep recognition, facial emotion identification, runtime recognition, and ocular tiredness assessment. Furthermore, eye state is a useful tool for developing HMIs and is extensively used in computer vision systems. Many people are prone to visual problems such as dry eye caused by computer use as a result of the recent technological advancements and the fact that computers are now a part of our daily lives. Computer vision syndromes (CVS) are a group of symptoms that are caused by people's inability to modify their eye condition (e.g., blinking) while concentrating at digital screens for a prolonged period of time. For this function, eye state identification is critical in identifying a person's blink condition when they are facing a screen. Having a small number of blinks on a digital display has both positive and negative ramifications. While the positive effects of blinking are linked with focus distraction and perception on displays, the negative impacts are associated with human health and are alarming since they equate to an increase in the number of persons infected with CVS. In the field of computer vision, eye state identification has become highly important. It contributes significantly to the improvement of human-computer interface technology by enabling precise eye state and blink recognition. Furthermore, there has been a boom in awareness in eye state research since identifying eye state raises awareness in many disciplines.

Driver weariness can also be determined by the condition of one's eyes. A number of methods are used to identify driver weariness, along with the observation of regulated equipment, physiological indications, and behaviors. Monitoring programmable devices is a non-invasive strategy with poor dependability because to the strong dependency on driver abilities and road quality. Controllable device screening requires the driver to connect data measuring devices to his body, making it nearly difficult to see these physiological indications. In behavioral and computer vision measurements, ocular features such as the degree of eye movement and the frequency of blinks are employed to identify weariness. Driver sleepiness is one of the most common causes of catastrophic vehicle accidents (insomnia, fatigue, inattention, and so on). Identifying driver sleepiness might be a key aspect of future autonomous cars. Tiredness in drivers may be detected using a variety of approaches and classified into three types: physiological, vehicle-based, and behavioral.

Physiological measures include electrocardiograms (ECG), electroencephalograms (EEG), and electrooculograms (EOG) collected via responsive electrodes or electronic devices worn by the driver. Physiological measures, while on the other hand, are rarely used since they obstruct the driver. Monitoring the vehicle's-controlled machinery (steering wheel, lane monitoring, and braking regulations) relies heavily on the driver's talents and road circumstances. This is another low-accuracy non-invasive sleepiness detection approach. Because behavioral views focus on the person instead of the resource, they are much more reliable than physiological and tool-based approaches. They rely on computer vision systems to detect weariness by analyzing the driver's behavior, facial expression, eye state, and blink condition using video-

recorded visual signals. Because of their lack of invasiveness and focus on the driver, behavioral approaches have lately gained appeal. Recent advances in fields such as face and eye identification and tracking, machine learning feature extraction, and deep learning have resulted in significant gain in eye state recognition. Nonetheless, because eye state recognition involves so many properties, it is continually changing on a daily basis. Early eye state recognition research concentrated on three circumstances: feature-based, motion-based, and appearance-based. In feature-based approaches, geometric features and gray-level patterns are utilized. Movement-based techniques are focused on the characteristics of eyelid movement. The tissue components of the eye are examined. in approaches based on physical appearance The outcomes of studies imply that view-based tactics outperform other strategies.

However, environmental factors play an important role in effectively diagnosing the eye problem. Various both internal and external components, such as illumination, light angle, head position, and image quality, can have a significant influence on the look and shape of the eyes, making exact quantification of the eye condition difficult. Because the actual world is loud and new circumstances are unexpectedly uncontrolled. In recent papers on eye state identification, machine learning methods such as AdaBoost and support vector machine (SVM) were proposed to improve the efficiency of recognition systems in unexpected (uncertain) circumstances. However, human techniques for feature extraction must be utilized in addition to machine learning approaches to recover the features. Moreover, because hand-crafted feature extraction methodologies demand a substantial amount of compute, the resulting systems are not only sluggish, but also require a significant amount of skill and experience.

Dong et al. [8] used Random Forest, Random Ferns, and Random Trees. And SVM algorithms for categorizing feature sets provided by different feature extraction methods for ocular state definition. They stated that the histogram-oriented gradient (HOG) was less affected by the noise effect for classification purposes, and their technique had a success rate of up to 93%. Pauly and sankar used low-resolution eye pictures to identify blinking. Sankar [9] used a variety of features (mean intensity, Fisher faces, and HOG feature) as well as classifiers like SVM and artificial neural network (ANN). The features learned by the HOG outperformed all other methods in the study when utilized with the SVM classifier, based on the comparative results of the five distinct methods employed in the research. Zhao et al. [10] introduced a deep integrated neural network that relied on eye area classification based on actionable intelligence. They tried many configurations by varying the training types in this integrated neural network and claimed that it produces the best results, allowing them to enhance the capacity to categories in tiny datasets by integrating transfer learning with data augmentation.

Deep learning technique advancement, as well as recent advances in artificial intelligence, has encouraged the creation of new approaches and concepts in picture classification. Because of CNN's excellent performance in image classification, one of the sub-branches of machine learning has had a significant impact in so many image-based technologies. As a result, instead of the hand-crafted feature extraction methodologies used in previous research, the use of deep learning-based, specifically transfer learning-based techniques in eye state recognition has emerged as an intriguing capacity in terms of both accuracy and efficiency.

Transfer learning is a method of learning. responding with minor differences across datasets by using the information gained by a neural network from one job to the next additional task for independent learning The network architecture is very essential in efficiency and speed of a deep network for picture classification In this research, we investigate the various architecture approaches and modifications presented in the GoogLeNet and inception networks. These versions are evaluated in terms of computing efficiency and the network characteristics and performances are compared using the ImageNet 2012 dataset, as well as a critical assessment of inception networks. To achieve this, global average pooling and dropout were utilized. Avoid overfitting. This resulted in an error in. lowered by 0.6% when compared to the scenario when the final layer was removed used as a completely linked layer The use of average pooling The inception layer is then followed by a 1X1 convolution. Inception-v4 displays a strong effect to that offered by Inception-ResNet variations' shortcut connectivity

Architecture	Top 5 Error
Inception-Resnet-V2	4.9
Inception-V4	5.0
Inception-Resnet-V1	5.5
Inception-V3	5.6
Inception-V2 Factorized 7x7	5.8
Inception-V2 Label Smoothing	6.1
Inception-V2 RMSProp	6.3
BN- Inception	7.8
GoogleNet	7.89

Table 1 TOP 5 Error

Several models were created using the GoogLeNet and ResNet designs, some advancements in these models were implemented, such as batch normalization [8], Smoothing of labels, inclusion of average pooling layers in the inception and reduction layers, as well as improved training Approaches were employed. All of these performances Models were evaluated using the ILSVRC 2012 classification. As seen in the table, the dataset has a top 5% error. The table displays the performance of GoogLeNet as well as the performance of momentum, RMS prop, and label in Inception-v2 smoothing. Application of these systems in conjunction with the use of Inception-v2 performed better after factoring. by 26.49% as compared to the GoogLeNet model Further, batch normalization of the auxiliary classifier in the network Inception-v2, i.e. The Inception-v3 network was supplied. 5.6% greater productivity. Inception-ResNet-v1 as well as Inception-ResNet-v2 was built to increase performance without deterioration in the deep layer design, which resulted in an even lower error rate.

2. METHODOLOGY

Using an eye state dataset, a DCNN-based approach is used. Was intended for automated identification of eye condition (open/closed). or (closed) in this study.

1. Resizing the image of close eyes and open eyes data for the CNN model divide data in training data and test data.
2. Train the data for the pre train CNN model for eyes state data to adjusting hyperparameter
3. Measuring the performance of CNN model to evaluating accuracy for reserved Teste data
4. Comparing both the model for eye state recognition

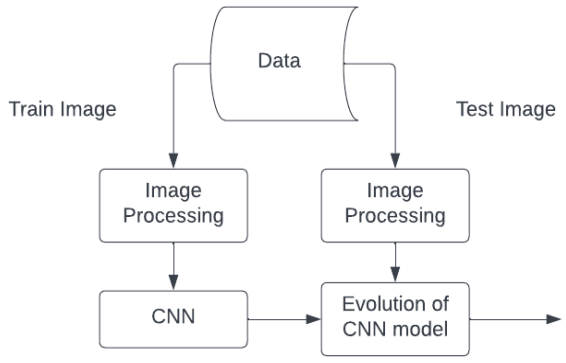
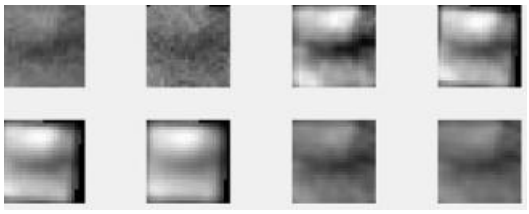


Fig 1 – Flow chart of Method

2.1 Data Collection

The photos were collected from the MRL dataset and were shot in typical lighting and brightness circumstances; thus, they were comparable to real-world scenarios. The dataset was obtained under tough circumstances caused by individual variances and other environmental conditions such as light, blur, and darkness, all of which are known in the actual world. This dataset was used to identify eye status or distinguish between open and closed eyes. Based on two criteria, the dataset was categorized into two parts (training and testing) (open and closed-eye images).

This dataset's open and closed eye photos are low-resolution, 24 * 24 pixels in size, and are also openly available. In this data set total 84,898 images available. we collected the data of 37 different persons (33 men and 4 women). t this moment, the dataset contains the images captured by three different sensors (Intel RealSense RS 300 sensor with 640 x 480 resolution, IDS Imaging sensor with 1280 x 1024 resolution, and Aptina sensor with 752 x 480 resolution). The dataset is suitable for testing several features or trainable classifiers. In order to simplify the comparison of algorithms, the images are divided into several categories, which also make them suitable for training and testing classifiers.



Close Eyes

Open Eyes

Fig 2 - Open & Close Eyes

2.2 Image Processing

The image processing component depicted in Fig. 1 enables the pictures in the dataset to be created in line with the entry into the training of pre-trained CNN architectures. These architectures' input layer sizes were used to determine resizing. The training and test pictures in the eye state dataset were downsized for GoogleNet, ResNet18 [35], MobileNetv2 and ShuffleNet (224 224), AlexNet (227 227), and DarkNet19 (256 256) according to the input parameters of the pretrained CNN architectures . The prepared dataset of ocular states was ready As a result of this approach; the produced eye state dataset was ready for the training and testing phases. This procedure was applied to two datasets based on the input of each trained CNN model.

2.3 Transfer Learning

CNNs are designed to continue functioning with pictures, which sets them apart from other methods. As a consequence, a 2D or 3D picture is automatically evaluated for CNN input. Another distinguishing feature of CNN is that it heavily relies on convolutional processes, as demonstrated by the "convolutional" abbreviation in its name. A simple CNN structure consists of three layers: the convolution layer, the pooling layer, and the fully linked layer. Following the convolution layer, subsampling layers such as normalization, activation, and pooling are used. stratum of evolution square number grids makes up the convolution layer (kernels).

These cores use convolution with the layer's data to construct and maintain the feature map. In other words, the kernel processes the layer's input from left to right and bottom to top while extracting the feature map. The mathematical formulation of the convolution process, the convolution of a continuous function x and w (x ∗w) (a), is defined in all dimensions by the following equation:

$$(x * w)(a) = \int x(t)w(a - t)da$$

In this case, is R^n for any n≥1. Furthermore, the higher dimensional form replaces integral. In practice, however, the parameter t is expected to be discrete; hence discrete convolution is defined as shown in the following equation:

$$(x * w)(a) = \sum a\, x(t)w(t - a)$$

where x is the input, w is the kernel, and the output is the feature map when a goes overall values in the input space. The pooling layer is primarily used after the convolution layer. This layer's primary goal is to minimize picture size by merging certain portions of the image into a single value, and it also shows the image's features. Another approach used after the convolution layer is the activation function. By training the non-linear predictions bounds, this variable is used to include non-linearity into deep learning models.

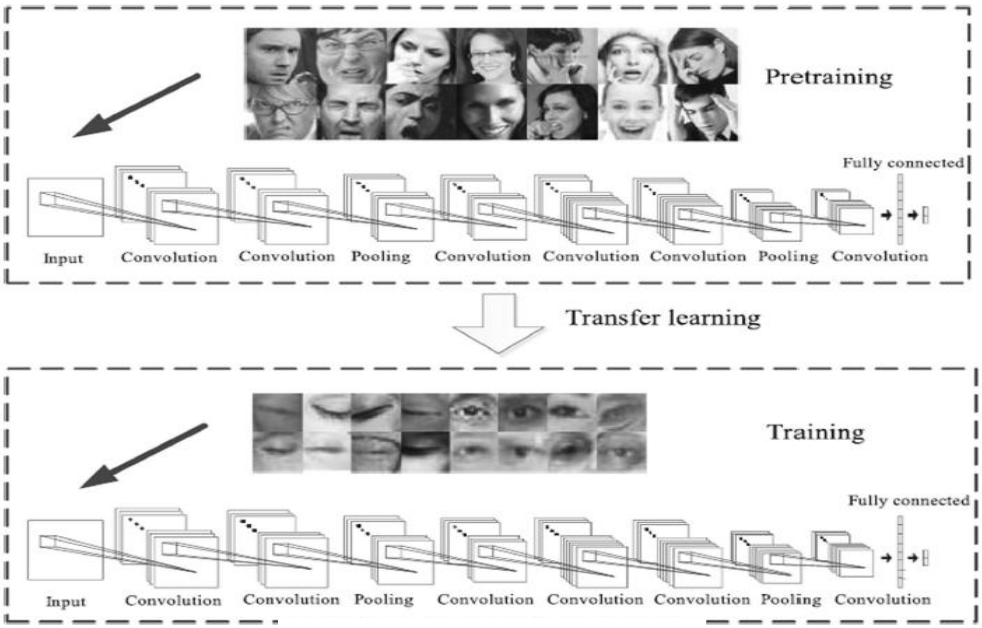


Fig – 3 Transfer Learning Process

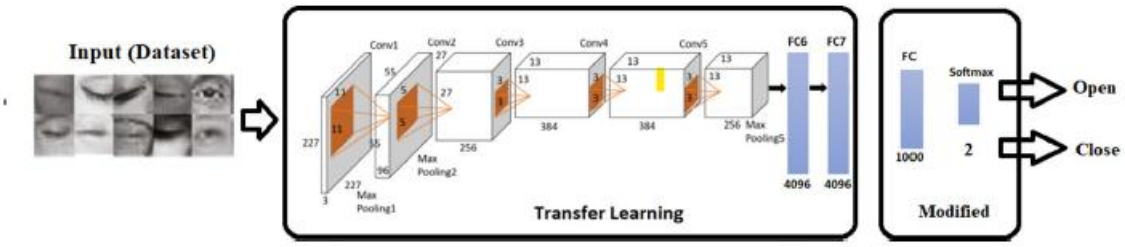


Fig – 4 Image Processing

2.4 Application

The problem of eye identification detection, which is addressed in this paper, is expected to meet real-world circumstances with varying degrees of difficulty. As a result, in addition to putting the proposed technique through its paces on the MRL datasets, It has also been put through its paces in a real - world situation. A flowchart was built in order for the suggested approach to classify the eye condition on the video. In addition, a film from a real-world event was used to show the efficacy of the suggested DCNN. The detection method retrieved the person's eye area from this video, and the acquired right and left eye regions were categorized in DCNN. In this video, DCNN trained on datasets was validated individually, and the most reliable eye state detection method was identified by comparing the results.

Because of its excellent detection rate and real-time performance, the Viola-Jones detector [47] was employed as the detection technique for real-time detection of faces and eye areas. This detector has been utilized in two methods to identify both the face and the ocular areas. The picture is first captured using the web camera, and then it is put into the face detector. The face detector identifies the matching facial region and outputs it. The result is then sent to the eye detector in the following step. The eye detector extracts two eye areas, right and left, from the facial image, and then feeds the extracted eye areas into the DCNN architecture for eye state detection. In the DCNN architecture, the two eye areas are classified individually. The classification produces two outputs, which are provided for assessment. All of these procedures are repeated until the pictures recovered from the movie are complete

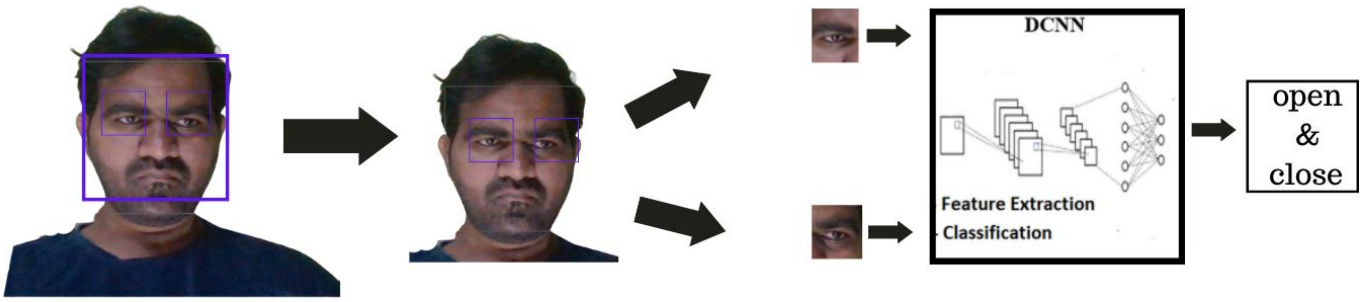


Fig – 5 Eye state identification application in real life scenario

3. CONCLUSION

Eye condition recognition has a wide range of applications, ranging from HMI systems to monitoring driver tiredness, dry eye, and computer vision syndrome caused by prolonged use of digital screens. Identification of eye status, either sensitively on or off, can pave the way for the development of a plethora of technology in this field. Many real-world events necessitate the identification of eye states. As a result, the proposed approach was put to the test by using a video from a realistic scenario. In this article, an eye state identification approach based on DCNN was demonstrated using the MRL datasets. In this model we use 10 epoch and got 94% accuracy, 1.4 loss and validation loss are 1.8 and validation accuracy is 92%. In the confusion matrices created as a result of these tests, it was seen that the proposed method trained with MRL showed the best performanceIn this data set we are using 82,000 images its include low light image as well as image of eyes with the glass so its give the good accuracy. The suggested approach also has been evaluated in a real-world setting, and the findings demonstrate that it performs well even under adverse conditions. For the future studies we can add some more images or combining another data and train the model for good accuracy

REFERENCES

[1] M. V. Sowmya Laxshmi, P. U, L. Chandana and S. N, "An Enhanced Driver Drowsiness Detection System using Transfer Learning," 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2021, pp. 1671-1678, doi: 10.1109/ICECA52323.2021.9676050.

[2] Qaisar Abbas, "HybridFatigue: A Real-time Driver Drowsiness Detection using Hybrid Features and Transfer Learning" International Journal of Advanced Computer Science and Applications (IJACSA), 11(1), 2020.

[3] Y. Xie, K. Chen, and Y. L. Murphey, "Real-time and Robust Driver Yawning Detection with Deep Neural Networks," Proceedings of the 2018 IEEE Symposium Series on Computational Intelligence, SSCI 2018, pp. 532–538, 2019

[4] Eyosiyas Tadesse, Weihua Sheng, Meiqin Liu,(2014) "Driver Drowsiness Detection through HMM based Dynamic Modeling." 2014 IEEE International Conference on Robotics Automation (ICRA) Hong Kong Convention and Exhibition Center

[5] Giri, Santosh Joshi, Basanta. (2019). "TRANSFER LEARNING BASED IMAGE VISUALIZATION USING CNN". Vol.11. 41-49. 10.5121/ijaia.2019.11404.

[6] Bakheet S, Al-Hamadi A. "A Framework for Instantaneous Driver Drowsiness Detection Based on Improved HOG Features and Naïve Bayesian Classification." Brain Sciences. 2021; 11(2):240

[7] Chirra, Venkata & Reddy, U. Srinivasulu & Kolli, Venkata Krishna Kolli. (2021). "Virtual facial expression recognition using deep CNN with ensemble learning." Journal of Ambient Intelligence and Humanized Computing. 12. 10.1007/s12652-020-02866-3.

[8] Dong, Y., Zhang, Y., Yue, J., Hu, Z.: Comparison of random forest, random ferns and support vector machine for eye state classification. Multimed. Tools Appl. **75**(19), 11763–11783 (2016)

[9] Pauly, L., Sankar, D.: Non-intrusive eye blink detection from low resolution images using HOG-SVM classifier. Int. J. Image Graph. Signal Process. **8**(10), 11 (2016)

[10] Zhao, L., Wang, Z., Zhang, G., Qi, Y., Wang, X.: Eye state recognition based on deep integrated neural network and transfer learning. Multimed. Tools Appl. **77**(15), 19415–19438 (2018)

BIOGRAPHIES

Jigar Sapkale received the B. E degree in Electronics and communication Engineering from Gujarat Technological University, in 2020.He is currently pursuing the M. Tech degree in Electronics and Telecommunication from Veermata Jijabai Technological Institute Mumbai, India.

Surendra Bhosale received B.E Electrical from Shivaji University, Kolhapur, Maharashtra and M.E. Degree in Electrical from the University of Mumbai, Maharashtra. Currently, He has Ph.D. Degree in Electrical Engineering from the University of Mumbai, India.. His teaching and research areas include Wireless Communications and Routing algorithms, Applications of Machine Learning and Deep Learning algorithms.