2021 18th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE). Mexico City, Mexico. November 10-12, 2021

Visual-based Real Time Driver Drowsiness Detection System Using CNN

Jonathan Flores-Monroy
Graduate Section of
ESIME Culhuacan
Instituto Politecnico Nacional
Mexico City, Mexico
0000-0002-2467-3600

Mariko Nakano-Miyatake
Graduate Section of
ESIME Culhuacan
Instituto Politecnico Nacional
Mexico City, Mexico
0000-0003-1346-7825,
mnakano@ipn.mx

Gabriel Sanchez-Perez
Graduate Section of
ESIME Culhuacan
Instituto Politecnico Nacional
Mexico City, Mexico
0000-0002-4735-205X

Hector Perez-Meana
Graduate Section of
ESIME Culhuacan
Instituto Politecnico Nacional
Mexico City, Mexico
0000-0002-7786-2050

***Abstract*—The traffic accident is one of the most frequent cause of death in the world; and an important cause of the traffic accident is the fatigue of the driver, who falls asleep during driving. To overcome this problem in this paper, we propose a real-time driver drowsiness detection system, in which the driver's face region is extracted and introduced into a specific designed shallow convolutional neural network (SS-CNN). The SS-CNN detects the state of driver drowsiness using "eye closure" or "eye open" state. To distinguish between the "eye closed" state caused by normal eye blinking and that caused by drowsiness, the proposed system analyzes consecutive results of the SS-CNN. If the system determines that driver falls asleep, an alarm rings to awake the driver in order to avoid a possible accident. The proposed SS-CNN provides an accuracy of 98.95%, which outperforms the previous works. In the experimental section, we provide several links in which real-time operations of the proposed system can be observed.**

***Keywords*—driver's drowsiness detection, Convolutional neural Networks, driver's fatigue, Real time implementation, Visual detection**

I. Introduction

The traffic accident is one of the most frequent causes of decease in the world, according to the World Health Organization (WHO) [1]. Approximately 23.5% of the car crushes in the United State are related to fatigue of the drivers who are in a drowse during driving [2]. In Mexico, 16,500 persons a year are killed in traffic accidents and its economic burden becomes approximately 150 billion Mexican pesos by year according to the report of the National Council for Prevention against Accidents (CONAPRA) [3].

To avoid traffic accidents caused by driver's drowsiness, in the expressway construction, a flat and straight road is intentionally avoided. However, these intentions are not sufficient, because the number of accidents caused by drowsy drivers has increased [4]. The Advanced Driver Assistance Systems (ADAS) provide several functionalities for assisting the driver, such as parking sensors, lane departure warning, traffic sign recognition, blind spot information system, and so on. However, until now in the conventional ADAS, the driver's fatigue detection is not included yet.

Considering the importance of the theme to avoid mortal accidents, several systems for driver's drowsiness detection are proposed in the literature [4-12]. The systems can be divided in three main categories depending on input signals used for drowsiness detection. The first category is driving pattern-based or vehicle behavior-based system, which uses the data obtained from the vehicle, such as velocity, acceleration, steering angle etc. [5, 6]. However, the signals obtained from vehicle are not related directly to the drowsiness of the driver. Additionally, these signals depend on driver's driving ability and type of vehicle. The second category uses driver's physiological signals [7-9], such as Electrooculogram (EOG), Electrocardiogram (ECG), Heart Rate and Electroencephalogram (EEG). Although these signals can be used effectively to detect the grade of driver's fatigue, to obtain these signals, the driver must put on the wearable sensors, which is invasive for driver and obstacle to drive.

The third category is visual-based system, in which video camera equipped inside of the vehicle takes frames of the diver's face, detecting eye blinking frequency, eyelid closure, yawning and facial expression to determine the driver's drowsiness [10-12]. Because this category of systems is not invasive for drivers and does not depend on the type of vehicle and driver's driving ability, researchers proposed several systems. Zhao et al. [10] proposed a visual features-based driver fatigue detection, in which Single-Shot Detector (SSD) is used to detect the face of driver and VGG16 Convolutional Neural Networks (CNN) is used to detect the video frames with closed eyelids. Finally, using PERCLOS values [13], which is rate of frames with closed eyelids over the whole frames, determine if the driver is drowsy or not. The system provides accuracy of 91.88% using NTHU-Drowsy Driver Detection (NTHU-DDD) data set [14]. The principal disadvantage of this system is the difficulty to implement the system in real time with a compact hardware, which can be equipped inside of the vehicle. The VGG16 with transfer-learning technique requires approximately 138 million of trainable data in the memory, which makes difficult its implementation. In [4], authors proposed driver's fatigue detection system, in which using EfficientDet-B0, eyes and mouth landmarks are detected and classified into "open" or "close" of eyes and mouth using both landmarks. The average accuracy of the fatigue detection using NTHU-DDD data set is 96.05% [4]. Although EfficientDet-B0 is considered as one of the fastest object detectors, it requires approximately 10 million of the trainable parameters, which makes difficult a compact implementation with low power consumption.

For a useful real-time driver's drowsiness detection system, we consider the following requirements: 1- Driver's drowsiness must be detected in real-time and the system alarms to awake the driver at just moment when the driver has a doze. 2- The memory and computational power requirements are as low as possible for compactness and low power consumption, which make possible a portable system for almost all types of vehicles. Taking in account the above requirements, in this paper we propose a visual-based real-time driver's drowsiness detector based on CNN. The proposed system detects driver's face using Viola & Jones algorithm [15] and detect drowsiness using a specific designed shallow CNN (SS-CNN). The proposed system provides 98.95% in accuracy using only 600K trainable parameters, which makes possible a compact implementation of the proposed system. Additionally, the proposed system triggers the alarm at the just moment when the system detects driver's drowsiness. The performance of the proposed system is evaluated using the NTHU-DDD data set and compared with the previous systems. The real performance of the proposed system can be observed in the links provided in the end of Section VI (Table IV).

The rest of the paper is organized as follows: in Section II, we provide some related works, which are real-time driver's drowsiness detection system. In Section III, the proposed system is described in detail. The experimental results and performance comparisons are provided in Section VI, finally we conclude our work in Section V.

II. Real-time drowsiness detection

In this section, we describe two related works. The purpose of both works is real-time detection of driver drowsiness [11, 12]. In the first work [11], Viola & Jones algorithm is applied to each video frame to detect face region. From the detected face region, 68 face landmarks are obtained using facial landmark detection algorithm [16], as shown by Fig.1. The 2D coordinates of the 68 face landmarks are extracted as feature vector, and it is introduced to the fully connected Artificial Neural Network (ANN) with three hidden layers to classify it into two classes: drowsy and no drowsy. The ReLU is used as activation function in all layers. Authors used the NTHU-DDD data set to evaluate their system, obtaining an accuracy of 80.92 %. This system is implemented in the mobile phone with android operating system.

In the second work proposed by Hashemi et al. [12], first the driver's face region is detected by Viola & Jones algorithm, then using 68 face landmarks (Fig. 1), one of the eyes regions (either left or right eye closer to video camera) is cropped. The cropped eye region is introduced to the trained Fully Designed Neural Network (FD-NN), which is composed by one Conv2D layer with 32 kernels, a Maxpooling2D with size 2×2, a Dropout layer with 25% dropout rate, Fully Connected layer and finally sigmoidal function that determines if eye region belongs to eye-closure class or eye-open class. The FD-NN is trained using ZJU Eyeblink dataset [17], which is set of gray-scale images of left or right open-eyes and closure-eyes with size of 24×24. In [12], authors reported only the classification performance of eye state (open or close) using ZJU test set as well as prediction time in this process. The overall performance composed by face detection, closed eye detection and decision-making process, if driver is in drowsy or not, is not reported.

Fig. 1. The 68 face landmarks used frequently for facial expression and eye or mouth detection.

III. Proposed Algorithm

The proposed real-time driver's drowsiness detector system is composed by the several stages, which are the face detection stage, face analysis stage based on a shallow CNN, which is a specific designed CNN (SS-CNN), and the stage of consecutive results analysis. According to the result of analysis, the system determines if alarm is rung or not. The block diagram of the proposed system is given by Fig. 2.

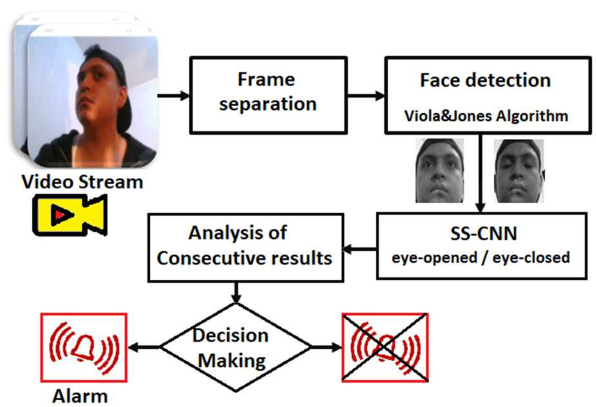


Fig. 2. Proposed real-time driver's doze detection system.

In the face detection stage, we used the Viola & Jones face detector [17], which extracts Haar features using integral image and the Adaboost classifier is applied in the cascading form to discard rapidly non-face object. During last few years, several CNN-based face detectors have been proposed, such as "Multi-Task Cascaded Convolutional Neural Network" (MTCNN) [18] and YOLO-Face [19], as well as Mediapipe-Face [20]. The last one is adapted for mobile devices. In general scenarios where there are multiple persons and lateral faces as well as frontal faces must be detected, the CNN-based face detectors present better detection performance compared with the Viola & Jones algorithm. However, in the real-time driver's drowsiness detection task, the video camera captures only a driver, and the driver looks straight ahead during driving, we decided to use the Viola & Jones algorithm due to its lowest memory requirement.

In Section IV-B, we compared the Viola & Jones algorithm with the Mediapipe-Face algorithm [20], taking in account of the memory requirement, as well as CPU and GPU occupancy rates.

Our specific designed shallow CNN (SS-CNN) used to detect if the driver with his/her eyes opened or closed, and the analysis stage of consecutive results of SS-CNN to determine the doze of driver are explained in the following subsections.

A. SS-CNN

We designed a shallow CNN to classify the face region into the face with eyes closed and the face with eyes opened. The input image for the SS-CNN is the gray-scale face region detected by the Viola & Jones algorithm. The important factors considered for its design are the number of trainable parameters of the CNN, which must be as small as possible, and the classification accuracy, which must be as high as possible. We evaluated several configurations of shallow CNN, varying number of Conv2D layers, number, and size of kernels in each Conv2D layer to select the best one that satisfies the above-mentioned requirements. The Table I shows the selected configuration. The number of the trainable parameters of the SS-CNN is approximately 600K, which is small enough to realize operations in real-time using a compact GPU-system.

TABLE I. PROPOSED SS-CNN STRUCTURE

Layer name	*Size of feature map*	*Number of parameters*
Conv2D	148 × 148 × 16	448
MaxPooling2D (2×2)	74 × 74 × 16	0
Conv2D	72 × 72 × 32	4640
MaxPooling2D (2×2)	36 × 36 × 32	0
Conv2D	34 × 34 × 16	4624
MaxPooling2D (2×2)	17 × 17 × 16	0
Flatten	4624	0
FC	128	592000
FC	1	129

We used ReLU activation function, which is expressed by $\phi(q) = \max(0, q)$, for all Conv2D layers and the first Fully connected (FC) layer. And sigmoid activation function, which is expressed by $\phi(q) = \frac{1}{1+e^{-q}}$, for the last FC layer to determine "eye closed" or "eye opened".

In the training stage of the SS-CNN, we used the Adaptive Moment Estimation (Adam) optimizer [21] to minimize binary cross-entropy loss function. For the SS-CNN the Adam optimizer shows a better performance compared with other optimizers. The Adam optimizer and the loss function are given by (1) and (2).

$$w_{t+1} = w_t - \alpha \frac{m_t}{\sqrt{v_t}+eps} \ , \tag{1}$$

where $m_{t+1} = \beta_1 m_t + (1-\beta_1)(w_t - w_{t-1})$, $v_{t+1} = \beta_2 v_t + (1-\beta_2)G_t$, $G_t = (w_t - w_{t-1})^2$ and *eps* is machine epsilon to avoid division by zero. We used $\beta_1 = 0.9, \beta_2 = 0.999$ and learning rate α is equal to 0.001.

$$loss = -\frac{1}{n}\sum_{i=1}^{n}\left(y_i log\hat{y}_i + (1-y_i)log(1-\hat{y}_i)\right), \tag{2}$$

where y_i and $\hat{y}_i$ are real and desired output, and n is number of batch size.

The SS-CNN is trained using the National Tsing Hua University Driver Drowsiness Detection (NTHU-DDD) datasets [14], which is describe in detail in Section IV.

B. Analysis for driver's drowsiness detection

The eye blinking is a normal behavior independent on driver's drowsiness and it forms a cycle composed by closing, closed, early opening and late opening. The duration of whole cycle is approximately 200-400 ms and the mean duration of eye closed state is 40-80 ms [22]. The principal factors of the variation are age and physical condition of each person. Considering the eye-closed state in the normal blinking and the frame rate of the Webcam, which is usually 30 fps, we decide that more than 4 consecutive frames with eye-closed state is driver drowsiness, which the driver is in a drowse. The Fig. 3 shows this process.

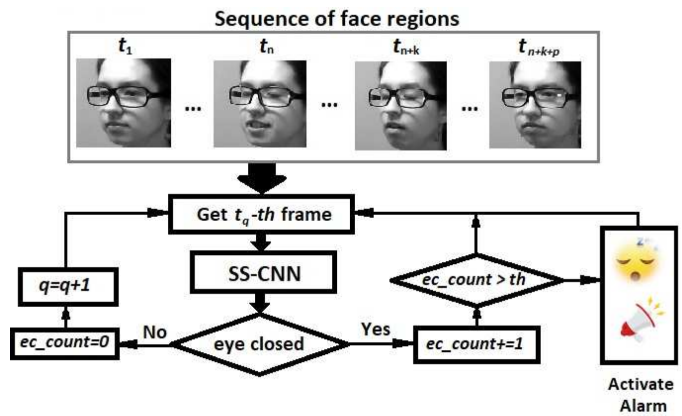


Fig. 3. Analysis of sequence of face region using SS-CNN and decision making

In Fig. 3, we used eye closed counter "ec_count" to detect the driver drowsiness. If driver's eyes continue closed during 4 frames (*th*=4), the driver falls asleep, according to the report [22] and then the alarm rings to wake up the driver in order to avoid a possible accident. The input sequence of face regions are generated by the Viola & Jones algorithm, as mentioned before.

IV. EXPERIMENTAL RESULTS

The proposed SS-CNN is trained using NTHU-DDD data sets and applied to video sequences taken in our laboratory to evaluate its performance considering real-time operability.

A. Training set and test set

We used the NTHU-DDD dataset [14] to the train proposed SS-CNN and evaluate the global performance of the proposed system under laboratory environment and real driving environment. The training set of the NTHU-DDD contains 18 subjects composed by males and females with different ethnicities. There are several videos per subject, varying types of illumination: visible right and infrared (IR) and accessories such as normal glasses, sunglasses and without glasses. We used only four types of videos, which are without glasses under visible light, without glasses under IR, with glasses under visible light and with glasses under IR light. The video sequences with sunglasses are not considered, because the proposed system determines driver drowsiness using “eye closed” or “eye opened”, it is difficult to observe the state of eyes through sunglasses even by human. All video frames have size of 640×480 pixels.

To adapt the NTHU-DDD dataset for our objective, which is detect in real time the state of the driver who falls asleep at the wheel and timely triggers the alarm to wake up the driver to avoid a possible accident. We labeled each frame of the video sequence as “eye closed” or “eye opened”. In total 4,620 face images are labeled, in which 2,310 images with “eye closed” and 2,310 images with “eye opened”. Fig. 4 shows some face regions, which are extracted using Viola & Jones algorithm and labeled as “eye closed” and “eye opened”. In the first row of the figure, the face images are labeled as “eye opened”, and the images in the second row are labeled as “eye closed”.

Fig. 4. Labeled dataset. Images in the first row are labeled as “eye opened” and images in the secod row are labeled as “eye closed”.

B. Face Detection Algorithms

To select an appropriate face detection algorithm for our purpose, we evaluated YOLO-Face [19], Viola & Jones face detector [17] and Mediapipe-Face [20]. The implementation of YOLO-Face in the compact GPU system, such as Jetson Nano, resulted unstable due to memory limitation and/or GPU limitation. Therefore, we compared only Mediapipe-Face [20] and Viola & Jones algorithm [17] to select most appropriate one. Table II shows a comparison of both algorithms from resource requirements and operation speed points of view.

TABLE II. PERFORMANCE COMPARISON BETWEEN VIOLA & JONES ALGORITHM AND MEDIAPIPE-FACE

Algorithm	CPU occupancy (%)	GPU occupancy (%)	Required RAM (GB)	Operation Speed (FPS)
Viola & Jones	14-23	0-10	6.9	14-29
Mediapipe-Face	14-25	3-8	7.5	4-8

We can observe in the table that the Viola & Jones algorithm requieres fewer memory space, attaining faster operation speed than Mediapipe-Face algorithm. Therefore, we selected the Viola & Jones algorithm to perform the face detection task.

C. Detection Accuracy and Comparison

To train the SS-CNN, 90% of 4,620 face images are used, and the rest 462 images (10% of total images) are used for evaluation. Fig. 5 shows the behavior of loss value and accuracy of the training and validation sets. From this figure, we can observe the good training behavior without overfitting.

After training, we evaluate the proposed SS-CNN performance with test set, and obtained total accuracy 98.95%, which is very good performance considering a large variation in illumination, the use or not of glasses and ethnicities in the face images. The accuracy obtained for face images with glasses is dropped down approximately 2% compared with that obtained for face images without glasses under both visible light and IR images. The infrared images provide very similar accuracy compared with the visible light images.

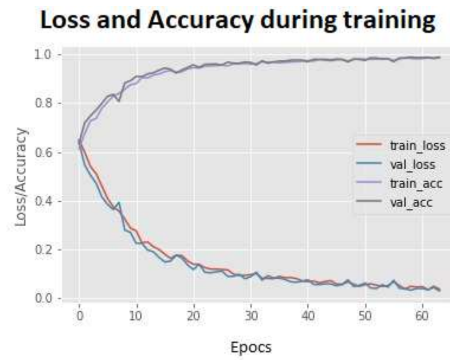


Fig. 5. Good training behavior without overfitting.

The performance of proposed SS-CNN is compared with the previous works that use the same dataset NTHU-DDD. Table III shows the performance comparison. From the table, we can observe that the proposed system outperforms the previously proposed real-time driver drowsiness system.

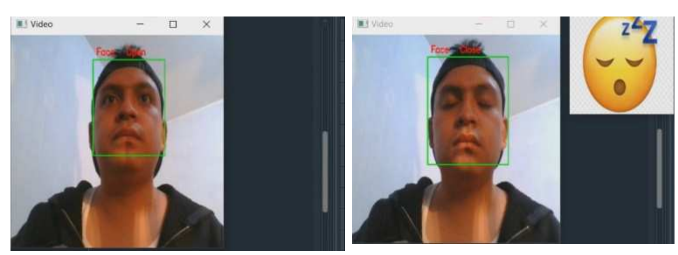
TABLE III. PERFORMANCE COMPARISON AMONG SEVERAL REAL-TIME DRIVER DROWSINESS DETECTION SYSTEMS

Works	Year	Accuracy
Jabbar [11]	2018	80.93%
Hashemi [12]	2020	98.15%
Ayachi [4]	2021	96.05%
Proposed	2021	**98.95%**

We used AMD Ryzen5 with 4,600 Hz and 16GB memory with 3,200Hz for evaluation in laboratory.

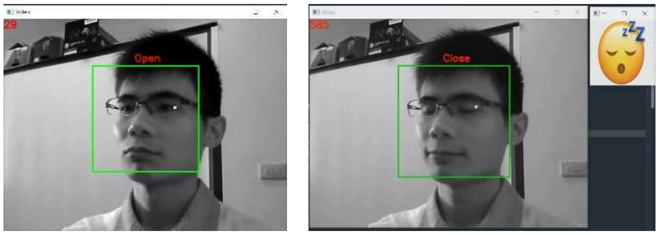
D. Real Time Performance

The whole proposed system shown in Fig. 2 is evaluated in laboratory environment and in the real driving situation. Fig. 6 and Fig.7 show the frames extracted from the real-time operation of the proposed system. In both figures, (a) shows the person’s eye is opened, while (b) shows that the person is asleep. The proposed system correctly detects eye state of the person; besides that, it can correctly distinguish between eye blinking and drowsiness.



(a) (b)

Fig. 6. The real-time operation performance of the proposed system.



(a) (b)

Fig. 7. The real-time operation performance of the proposed system, applying to one of the video sequence provided by NTHU-DDD dataset.

We also evaluated the performance of the proposed system under real driving situation. To evaluate the proposed system in a real condition, it was installed inside of a car using the hardware system with the following descriptions: Nvidia Jetson Nano with Quad-core ARM A57 @ 1.43 GHz and GPU 128-core Maxwell with 4 GB memory, and Logitech Webcam with the 720p video quality at 30 fps and CSI IMX219-77IR infrared camera with a 720p video quality at 30 fps. Although Jetson Nano has capacity to operate in real time for 4K quality videos at 60 fps [24], we selected these two Webcams for our implementation. Evaluation results can be observed in the video links described in Table IV.

TABLE IV. LINK TO THE VIDEOS WITH BRIEF DESCRIPTION

Link address	Description
https://youtu.be/9rFOXpMPiGA	Real-time demonstration in laboratory
https://youtu.be/Sjbj3Ae2q60	Real-time demonstration using NTHU-DDD dataset
https://youtu.be/f0K_lFp1QOc	Real-time demonstration under real conditions in a car

V. CONCLUSIONS

In this paper, we proposed a real-time driver drowsiness detection system, in which from each video frame, the face region is extracted using Viola & Jones algorithm. Then the face region is introduced to the Specific designed Shallow CNN (SS-CNN) trained by the NTHU-DDD dataset. The SS-CNN provides the information about the face, whose eye is closed or opened. Using this information provided by the SS-CNN, the proposed system analyzes if the "eye closed" is caused by normal eye blinking or drowsiness of driver, considering the duration of "eye closed" state. If "eye closed" continues more than 4 consecutive frames, the system determines that driver is asleep and triggers the alarm to wake up the driver in order to avoid a possible accident.

The drowsiness detection performance provided by the proposed system are compared with previously proposed driver drowsiness system. The comparison results show that the proposed system outperforms the previous works. The real-time performances of the proposed system under different situations also evaluated, which can be observed in the videos posted in the links provided in Table IV.

The driver drowsiness or fatigue can be measured using the frequency of yawning or some specific movement of head of the driver. As future work, we consider other signs of drowsiness or fatigue to combine with this proposed system.

REFERENCES

[1] World Health Organization: The top ten causes of death. [online]. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-deth.

[2] Drowsy Driving NHTSA reports. [online] https://www,nhtsa.gov/risku-driving/deowsy-driving.

[3] M. Páez and E. Abarca, "Tools for security in movements, Predictive models of driver's drowsiness" (Spanish) 2017, [online] https://imt.mx/resumen-boletines.html?IdArticulo=449&IdBoletin=168. [Accessed: 01- Jun- 2021].

[4] R. Ayachi, M. Afif, Y. Said and A. B. Abdelali, "Drivers fatigue detection using EfficientDet in Advanced Driver Assistance System," in Proc. Int. Multi-conference on Systems, Signal & Devices, 2021, pp. 738-742.

[5] J. Wang, S. Zhu and Y. Gong, "Driving safety monitoring using semisupervised learning on time series data," IEEE Trans. Intell. Transp. Syst. Vol. 11, no. 3, pp. 728-737, Sep. 2010.

[6] B.-F. Wu, Y.-H. Chen, C.-H. Yeh and Y.-f- Li, "reasoning-based framework for driving safety monitoring using driving event recognition," IEEE Trans. Intell. Transp. Syst, vol. 14, no. 3, pp. 1231-1241, sep. 2013

[7] A. K. kokonogi, E. M. Michail, I. C. Chouvarda and N. M. Maglaveras, "A study of heart rate and brain system complexity and their interaction in sleep-deprived subjects," in Proc. Comput. Cardiol., Sep. 2008, pp. 969-971.

[8] V. Vidente, P. Laguna, A. Bartra and R. Bailón, "Drowsiness detection using heart rate variability", Med. Biol. Eng. Comput. vol. 54, no. 6, pp. 827-937, 2016.

[9] C. Zhang, H. Wang and R,.Fu, "Automated detection of driver fatigue based on entropy and complexity measures, " IEEE Trans. Intell. Transp. Syst, vol. 15, no. 1, pp. 168-177, 2014.

[10] G. Zhao, Y. He, H. Yang and Y. Tao, "Research on fatigue detection based on visual features," IET Image Processing, pp. 1-10, 2020.

[11] R. Jabbar, K. Al-Khalifa, M. Kharbeche, W. Alhajyaseen, M. Jafari and S. Jiang, "Real-time Driver Drowsiness Detection for Android Application using Deep Neural Networks Techniques," Procedia Computer Scence vol. 130, pp. 400-407, 2018.

[12] M. Hashemi, A. Mirrashid and A. B. Shirazi, "Driver Safety Development: Real-time driver drowsiness detection system based on convolutional neural networks," Springer Nature Computer Scioence, 1:289, 2020.

[13] M. Flores, J. Armingo and A. de la Escalera, "A: Real-time K. Elissa, "A. Real-time warning system for driver drowsiness detection using visual information, " J. Intell. Rob. Syst. C. vol 69, no. 2, pp. 103-125, 2010.

[14] C. H. Weng, Y. H. Lai and S. H. Lai, "Driver drowsiness detection via a hierarchical temporal deep belief network", Proc Asian Conference on Computer Vision, pp. 117-133, 2016.

[15] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1867-1874, 2014.

[16] G. Pan, L. Sun, Z. Wu and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcamera," In Proc. IEEE Int. Conf. on Computer Vision, pp. 1-8, 2007.

[17] P. Viola and M. Jones, "Robust real-time object detection", In Proc. 2nd Int. workshop on Statistical and Computation Theories of Vision – Modeling, Learning, Computing and Sampling, p. 25, 2001.

[18] K. Zhang, Z. Zhang, Z. Li and Y. Qiao, "Joint face detection alignment using multi-task cascaded convolutional networks," IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499-1503, 2016

[19] W. Chen, H. Huang, S. Peng, C. Zhou and C. Zhang, "YOLO-face: a real-time face detector, " The Visual Computer, vol. 37, pp. 805-813, 2021.

[20] V. Bazarevsky, Y. Kartynnik, A. Vakunov, K. Raveendran, M. Grundmann, "BlazeFace: Sub-millisecond Neural Face Detection on Mobile GPUs," URL: https://arxiv.org/abs/1907.05047.

[21] D. P. Kingma and J. Ba. "Adam: A Method for Stochastic Optimization". In:CoRR abs/1412.6980 (2014). URL: http://arxiv.org/abs/412.6980

[22] K-A Kwon, R. Shipley, M. Edirisinghe, D. G. Ezra, G. Rose, S. M. Best and R. E. Cameron, "High-speed camera characterization of voluntary eye blinking kinematics," J. of the Royal Society Interface, pp. 1-6, 2013.