Received November 26, 2019, accepted December 5, 2019, date of publication December 10, 2019, date of current version December 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2958667

A Real-time Driving Drowsiness Detection Algorithm With Individual Differences Consideration

FENG YOU[1], XIAOLONG LI[1], YUNBO GONG[1], HAIWEI WANG[2], AND HONGYI LI[3,4]

[1]School of Civil Engineering and Transportation, South China University of Technology, Guangzhou 510640, China
[2]School of Transportation and Economic Management, Guangdong Communication Polytechnic, Guangzhou 510650, China
[3]Xinjiang Quality of Products Supervision and Inspection Institute of Technology, Urumqi 830011, China
[4]Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Urumqi 830011, China

Corresponding authors: Haiwei Wang (whw2046@126.com) and Hongyi Li (lihy@ms.xjb.ac.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51808151 and Grant 51408237, in part by the Guangdong Natural Science Foundation 2020, in part by the Guangdong Provincial Public Welfare Research and Capacity Building Special Project under Grant 2016A020223002, in part by the Guangdong Provincial Science and Technology Plan Project under Grant 2017A040405021, and in part by the Fundamental Research Funds for Guangdong Communication Polytechnic under Grant 20181014.

ABSTRACT The research work about driving drowsiness detection algorithm has great significance to improve traffic safety. Presently, there are many fruits and literature about driving drowsiness detection method. However, most of them are devoted to find a universal drowsiness detection method, while ignore the individual driver differences. This paper proposes a real-time driving drowsiness detection algorithm that considers the individual differences of driver. A deep cascaded convolutional neural network was constructed to detect the face region, which avoids the problem of poor accuracy caused by artificial feature extraction. Based on the Dlib toolkit, the landmarks of frontal driver facial in a frame are found. According to the eyes landmarks, a new parameter, called Eyes Aspect Ratio, is introduced to evaluate the drowsiness of driver in the current frame. Taking into account differences in size of driver's eyes, the proposed algorithm consists of two modules: offline training and online monitoring. In the first module, a unique fatigue state classifier, based on Support Vector Machines, was trained which taking the Eyes Aspect Ratio as input. Then, in the second module, the trained classifier is application to monitor the state of driver online. Because the fatigue driving state is gradually produced, a variable which calculated by number of drowsy frames per unit time is introduced to assess the drowsiness of driver. Through comparative experiments, we demonstrate this algorithm outperforms current driving drowsiness detection approaches in both accuracy and speed. In simulated driving applications, the proposed algorithm detects the drowsy state of driver quickly from $640 * 480$ resolution images at over 20fps and 94.80% accuracy. The research result can serve intelligent transportation system, ensure driver safety and reduce the losses caused by drowsy driving.

INDEX TERMS Traffic safety, driving drowsiness detection, CNN, individual differences, SVM.

I. INTRODUCTION

With the increasing number of vehicles all over the world, traffic accidents have become one of the primary reasons to cause human death. According to the report of the World Health Organization (WHO), traffic accidents are one of the top ten causes of human death in 2015 [1]. As we know, the driver, as the core of the road traffic system, is the most significant factor affecting road traffic safety. According to the record of National Sleep Foundation, about 32% of drivers have at least one drowsy driving experience per month [2]. About 100,000 accidents caused by drowsy driving and about 25% of traffic accidents involve drowsy driving every year [3]. Driving drowsiness refers to the behavior of driving skills declining objectively, due to the imbalance of physiological functions after driving continuously for a long time [4]. It may affect driving behavior and pose a serious safety threat to drivers and other traffic participants. To prevent sleepy driving, a series of laws and regulations are enacted all over the world. For example, the Road Traffic

The associate editor coordinating the review of this manuscript and approving it for publication was Amr Tolba.

Safety Law of China stipulates that the driver must not run a vehicle continuously for more than 4 hours [5]. By measuring driving time to judge the state of drivers, it can reduce the traffic accidents caused by driving drowsiness to some extent. However, this method cannot detect whether a driver is fatigued or not in real-time. With the development of information technology, detection system for driving drowsiness has become an alternative means to solve the problem. Therefore, research on the intelligent identification of drowsy driving has important realistic meanings.

In recent years, thanks to the non-invasive and low cost, methods of driving drowsiness detection based on driver behavior have become a research hotspot. However, the performance of algorithms might be limited by the technologies of face detection and fatigue assessing in complex and changing surroundings [6], [7]. Reference [8] uses a mask to obtain the face of driver and evaluate the driving state with the application of PERCLOS. Experiments show that the performance of the method is well good in ideal conditions. But the generalization performance of the method is affected by the fabrication of mask. The cascade face detector proposed by Viola and Jones [9] utilizes Haar-like feature and Adaboost to train cascaded classifier, which achieve good performance in face detection. [10] extracts the multi-scale feature of face using Gabor wavelet transform, and trains an Adaboost cascaded classifier to select a most recognizable features for driver drowsy state discrimination. However, quite a few works [11]–[13] indicate that the method based on Adaboost may degrade significantly in read-world application with larger visual variations. Reference [14] uses an active near-infrared light source to obtain the stable image of eyes and tracks it using finite state machine. The fuzzy system is applied to evaluate the state of driver. However, the method requires the installation of more sophisticated sensors, which increases the cost. With the development of artificial intelligence technology, the capable of computing has been greatly improved. The driving drowsiness detection algorithms based on deep learning are increasingly attracting attention. Reference [15] proposed a fatigue detection algorithm based on the Convolutional Neural Network (CNN). The algorithm classified human eyes and non-human eyes by training the first network, detected the position of the eye feature points with the second network, calculated the eye-opening degree according to the feature point position, and judged the fatigue state of driver by PERCLOS. Reference [16] introduced a detection method based on facial behavior analysis. It combined with Adaboost and kernel correlation filter (KCF) for face detection and tracking in facial image captured from infrared acquisition device. The cascading regression is exploited to locate the feature points and extracted the eye and mouth regions. Also, the convolutional neural network (CNN) is employed to recognize the status of eyes and mouth. However, the deep learning methods are still in the exploration stage for the driving drowsiness detection. A lot of works [17]–[21] were focus on exploring a universal algorithm based on deep learning technology while ignore the specific application sometimes.

Percentage of Eyelid Closure Over the Pupil Over Time(PERCLOS) [22] is one of the most popular parameter, which applies to driving drowsiness detection based on computer vision. It firstly assesses whether the driver's eyes are open or closed in the current frame according to the proportion of eyelid covering the pupil. Then, PERCLOS is calculated by the number of eyes closed frames over a pried of time. As we all know, the characteristics of human facial, especially the size of eyes, have certain differences, which introduces a new variable when it comes to PERCLOS. In the past literatures [23], PERCLOS-80(P80) was used to judge whether the eyes of driver are open or closed. That is, when the proportion of eyelid covering the pupil is over 80 percent, the eyes are identified as closed in the current frame. However, this method does not take into account the individual differences of driver, especially the differences of the size of eyes, which may cause misjudgment in practical applications.

In this paper, we propose a new algorithm to detect driving drowsiness, which considers the individual differences of the drivers. It consists of off-line training module and online monitoring module. Firstly, we design a deep cascaded convolutional neural network (DCCNN) to detect the face region from live video. Based on this, the eyes landmarks are obtained by the application of Dlib toolkit. A new parameter, Eyes Aspect Ratio (EAR), which calculated by the coordinates of landmarks, is introduced to identify the state of eyes (open or closed). In the offline training module, for a specific driver, there are two sets of the EAR, which represent eyes-open and eyes-closed, are obtained from the face detected by DCCNN. And a unique support vector machine (SVM) classifier is constructed which takes the two sets of data into the input. In the online monitoring module, the driver's face is detected by DCCNN from live video firstly. Then the landmarks of driver's face are obtained with the application of the Dlib toolkit as well. The Eyes Aspect Ratio can be calculated by the coordinates of the eyes. Finally, the state of eyes in the current frame can be classified by the unique SVM. Thanks to this offline training and online monitoring module, the performance of the algorithm can be notably improved.

The major contributions of this paper are summarized as follows:

1. We design a new deep cascaded convolutional neural network to detect the face of a driver, which effectively improve the performance.

2. We introduce a new parameter based on the Dlib toolkit to assess whether the eyes of a driver are open or closed.

3. Extensive experiments are conducted, to show the EAR is work in different drivers and the significant performance improved of the proposed approach in both accuracy and speed.

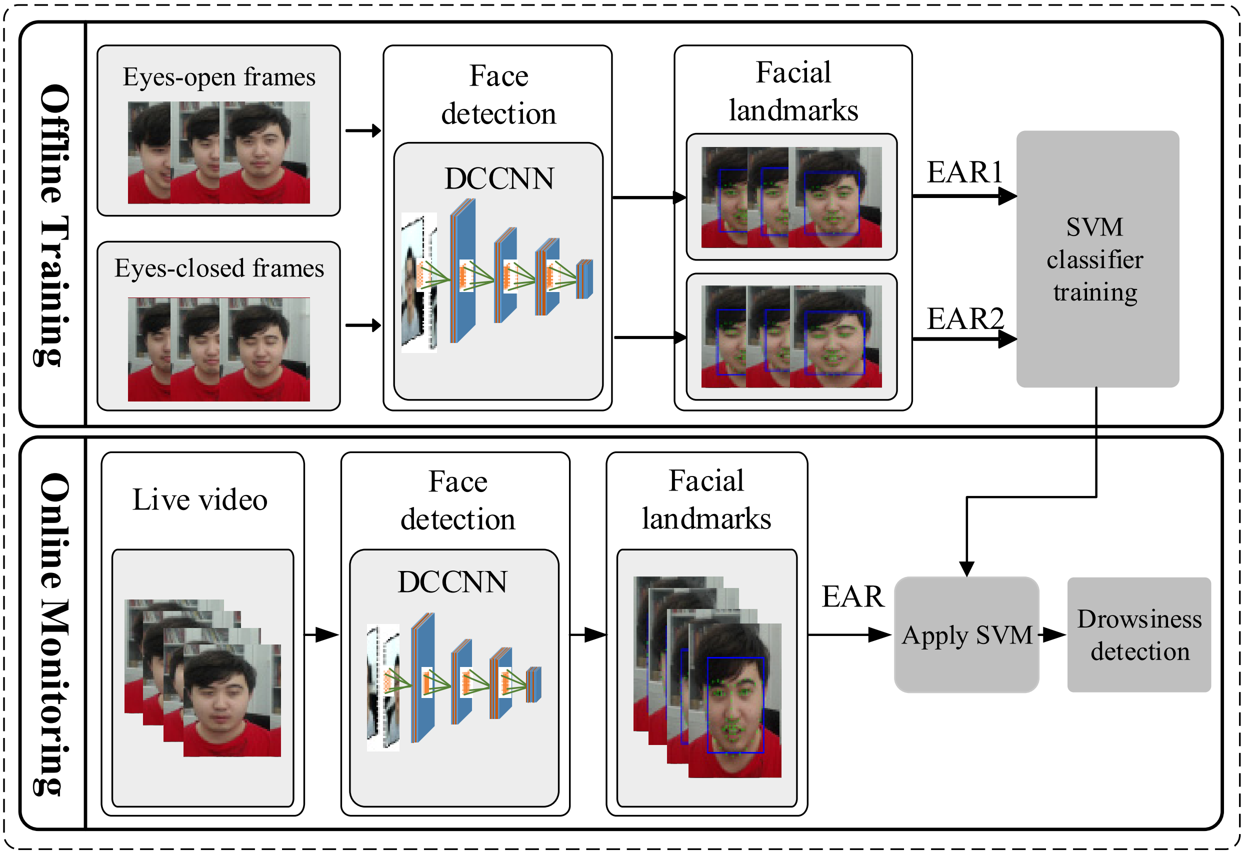


FIGURE 1. **Pipeline of proposed approach.**

4. Comparative experiments result show that the proposed algorithm, which with consideration of individual differences, is reasonable and more accurate.

The structure of this paper is organized as follows. In Section II, we describe the proposed methods in detail, including the deep learning model for face detection and facial landmarks obtaining. The core approaches, offline training and online monitoring, are introduced in this section as well. In Section III, we conduct extensive experiments to verify the rationality of the proposed algorithm in both accuracy and speed. The conclusion is in Section IV.

II. APPROACH

The overall pipeline of our approach is shown in Figure 1. The algorithm consists of the following two modules:

Offline Training: It is equivalent to an initialization process. As usually seen, the characteristics of a driver's face, especially the size of the eyes, have obvious differences. Thus, we firstly obtain two sets of data by asking the driver to open his/her eyes and close his/her eyes for a while on a simulator. For each image data, applying the deep cascaded convolutional neural network we designed called DCCNN to detect the face of drive in the current frame. Further, based on the Dlib toolkit, facial landmarks can be obtained. Then, two types of EAR, as shown in Figure 1, the EAR1 relates the eye-open frames and the EAR2 is the value calculated when the state of driver's eyes are closed, can be calculated. Finally, take the two types of the EAR as positive and negative samples, we training a unique SVM classifier for the specific driver to judge whether the eyes are open or closed.

Online Monitoring: It is a real-time module to detect driving drowsiness form live video. All frames are fed to DCCNN during driving to find the face of a driver. If the face is captured, the eyes landmarks can be obtained by the Dlib toolkit. Then, the unique SVM classifier of the specific driver, which trained in offline module, is applied to judge whether the eyes are open or closed in the current frame. At last, we assess whether the driver is sleepy or not according to the ratio of the number of sleepy frames and total frames in a period of time. Besides, if DCCNN does not detect a face, we judge the driver is drowsy in the current frame maybe because of the improper head posture.

A. DEEP CASCADED CONVOLUTIONAL NEURAL NETWORK

Face detection is one of the most important technologies for driving drowsiness detection based on computer vision. In practical application, driving drowsy detection system not only requires high accuracy, but also fast speed. As we know, deep learning methods, especially the convolutional neural network model, greatly improve the accuracy of image recognition. However, the complex network structure reduces the algorithm speed. In [24], a Multi-Task Cascaded Convolutional Networks called MTCNN have been designed

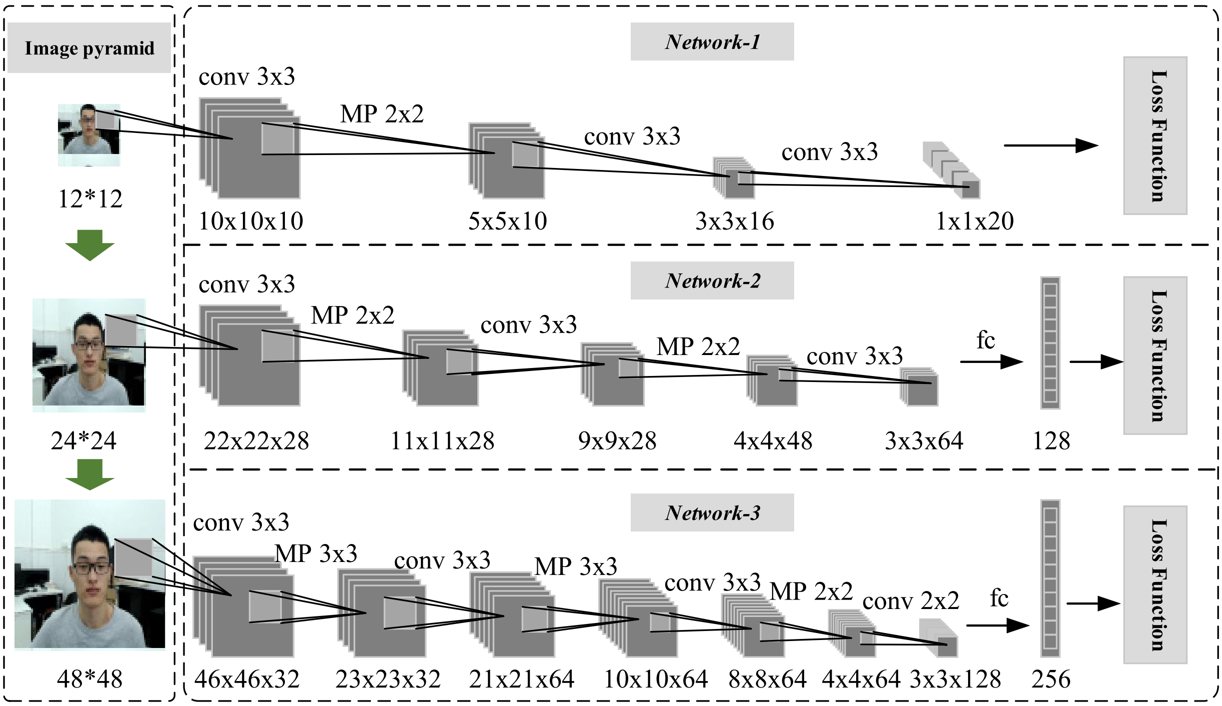


FIGURE 2. The architecture of DCCNN, where "conv" means convolutional layer, "MP" means max pooling layer and "fc" means fully connected layer.

for face detection. The architecture consists of three sub-networks, instead of a complex network. Each sub-network has less numbers of filters but more discrimination of them, which effectively improves the speed of the algorithm. However, we noticed its performance might be limited by the useless information about the five facial landmarks. Because we just pay attention to the information of the facial in video frames. Thus, we design a new convolutional neural network to detect the face of a driver. It is more efficient thanks to the lightweight architecture and removing five facial landmarks. The architecture of Deep Cascaded Convolutional Neural Network (DCCNN) is shown in Figure 2. In practical application, the proportion of the face in frames is uncertain. Thus, it is an effective approach to resize the original image into different scales. Similar to MTCNN, we build an image pyramid by resizing the original image to different scales, which is the input for three cascaded networks. Network-1 is a fully convolutional network [25], to obtain a large number of candidate windows and their bounding box regression vectors. Then we use the estimated bounding box regression vectors to calibrate the candidates. Finally, non-maximum suppression(NMS) [26] is employed to merge highly overlapped candidates. After Network-1, All candidates are fed to the next CNN, called network-2, which further reject a larger number of false candidates, performs calibration with bounding box regression, and NMS candidate merge. Network-3 is a CNN similar to network-2, but it has more convolutional layers and pooling layers so that it can describe the face in detail. The output of network-3 is the facial bounding box with high confidence. The relationship of the three sub-networks is cascaded. In DCCNN, the output of the previous sub-network acts as the input to the next sub-network. By building the image pyramid, the face of a driver can be detected accurate no matter how larger proportion of the face in frames.

In the training phase of DCCNN, the WIDER_FACE data set [27] and the AFLW data set [28] are used as the training data. Among them, the WIDER_FACE data set includes more than 30,000 pictures and 400,000 personal faces. It is the largest and most complex face detection public data set in the world. The AFLW face database includes 25,000 hand-labeled face images. The database is widely used in face recognition, face detection, face alignment and other aspects. In this paper, the WIDER_FACE and AFLW datasets are used to crop some face images and non-face images based on manually labeled face regions. A total of 190,000 face images, 600,000 partial face images and 900,000 non-face images are obtained.

The procedure of the training of the DCCNN is aiming to get the optimal model by adjusting the parameters dynamically. To express the influence of different parameters on the performance of the DCCNN, a loss function is introduced during training. The loss function is an index to measure the difference between the predictive output and the actually marked label, and there are various loss functions according to different tasks. The training process includes two tasks, respectively, face and non-face classification task and face region box bounding fitting task.

The first task, face and non-face is a classification task, so we apply the cross entropy [29] loss function for the training.

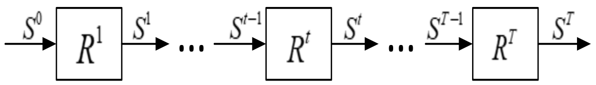


FIGURE 3. Diagram of CPR.

For any sample x_i, the cross-entropy loss function is:

$$L_i^1 = -(y_i^1 \log(p_i) + (1 - y_i^1)(1 - \log(p_i))) \tag{1}$$

where p_i is the predictive output of network, y_i^1 is the real label of x_i (face/non-face).

The second task is used to predict the coordinates of the face region box bounding, which belongs to the regression problem. Therefore, the Euclidean loss functions are applied for training. The loss function is shown in equation (2).

$$L_i^2 = \left\| p_i - y_i^2 \right\|_2^2 \tag{2}$$

where p_i is the coordinates of the network prediction face region, and y_i^2 is the real coordinate of the face region in the image.

Through the training of the deep cascaded convolutional neural network, the face of driver can be obtained accurate, which provides a stable face image for the following algorithms.

B. FACIAL LANDMARKS AND EAR

1) FACIAL LANDMARKS OBTAIN BASE ON CPR

In this paper, the driving drowsiness is monitored by the eyes state of driver. It is crucial to obtain the corner and shape of eyes. Cascaded Pose Regression(CPR) [30] is a regression algorithm to estimate the pose of object. Particularly, the pose of facial can be represented by feature points or landmarks. Therefore, we introduce the CPR algorithm to estimate the pose of driver by getting facial landmarks.

The diagram of CPR is shown in Figure 3. Given a S^0, which is an initial pose for the first weakly regressor. It will return a final estimated pose S^T after T iterations, and each of predicted pose is processed by different weakly regressor $R^i, i = 1, 2, \ldots, T$. Notice that the input of each weakly regressor depends on the output of previous one and each weakly regressor will extract the pose index feature x, which is crucial to express the pose of face. In the process of training, all weakly regressor automatically learn from the train samples, which were labeled in advance. The labeled samples are 2D face images that containing the landmarks of the feature points, that is, $S = \{x_1, y_1; x_2, y_2; \ldots; x_n, y_n\}$, where (x_i, y_i) is one of the coordinate of landmarks.

The training process of CPR is as follows:

Given a 2D face image I, from 1 to T do:

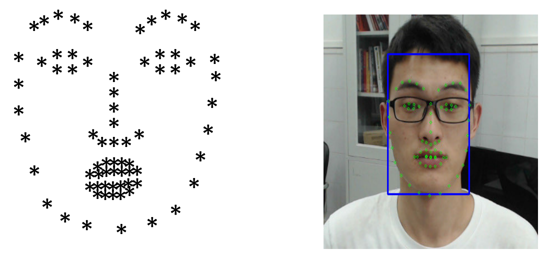
Calculate the pose index feature by:

$$x^t = h^t(S^{t-1}, I) \tag{3}$$

where h^t is a pose-index features as [30].

Evaluate regressor by:

$$S_\delta = R^t(x^t) \tag{4}$$



(a) 68 landmarks of face (b) facial landmarks obtain

FIGURE 4. Facial landmarks obtain based on Dlib.

FIGURE 5. Ellipse fitting. *Left:* original eye image, *middle*: segmented image, *right*: fitted ellipse.

Update pose S^t by:

$$S^t = S^{t-1} + S_\delta \tag{5}$$

Finally, the output of CPR is the estimated pose of face S^T.

Dlib [31] is an open source toolkit that includes machine learning algorithms and tools for creating complex algorithms to solve real-world problems. It is widely used in industrial and academia, including robotics, embedded devices, mobile phones and large high-performance computing environments. In Dlib toolkit, the pose of face is represented by 68 landmarks, as shown in Figure 4. The toolkit generates a regressor to obtain the facial landmarks of driver after training the CPR mentioned above, with input of 68 labeled landmarks.

2) EAR: A STABLE PARAMETER FOR EYES STATE EVALUATION

As mentioned, to a certain extent, the state of eyes indicates whether the driver is drowsy or not. Because there are significant differences about time of eyes closed between awake and drowsy. In [32], a method of ellipse fitting was proposed to describe the shape of pupil. As shown in Figure 5, the method segments the pupil with traditional image process firstly. Then, an ellipse is fitted with the white pixels, which represent the shape of eyes. Lastly, the ratio of the major and minor axes of the ellipse was used to evaluate the eyes state.

We noticed its performance might be limited by the following facts: (1) The pixel values are sensitive. Changeable environment is easy to make image segmentation to be worse. (2) In practical application, the pixel values between pupils and glasses are very close, which lead to false ellipse fitting. In this paper, we design a new more stable parameter based on Dlib toolkit to evaluate the state of driver's eyes. It is more stable and precise than ellipse fitting method thanks to avoiding the traditional image process.

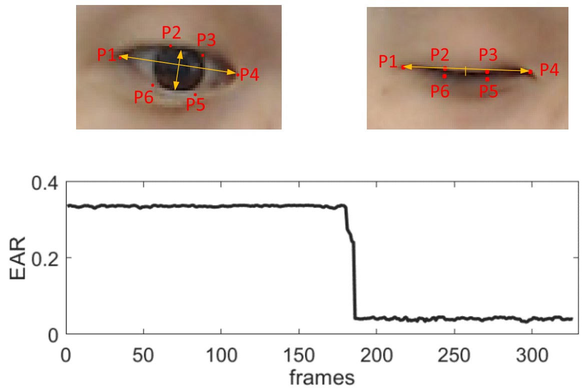


FIGURE 6. Eyes landmarks. ***Upper:*** the distribution of eyes landmarks has significant differences. ***Bottom:*** the values of EAR at open and closed state.

In section B-1), we have obtained the facial landmarks based on Dlib toolkit. As shown in Figure 6, for each eye, there are six points distributed around to locate the position of eye. The distribution of eyes landmarks has significant differences between open and closed state. In [33], Eye Aspect Ratio was application to record the blink frequency.

EAR can be computed according to the position of eyes landmarks by:

$$EAR = \frac{\|P_2 - P_6\| + \|P_3 - P_5\|}{2\,\|P_1 - P_4\|} \tag{6}$$

where $P_i,\ i = 1, 2, \ldots, 6$ is the coordinate of eyes landmarks.

As shown in Figure 6, when the eyes of driver are open, the EAR is over 0.2. In contrast, the EAR is less than 0.2. Thanks to the stable facial landmarks location model based on CPR, the new parameter, EAR, is much more robust than [32].

C. DROWSINESS ASSESSMENT MODEL

1) OFFLINE TRAINING

As mentioned, P80 uses a hard threshold (0.8) to judge whether the eyes of driver are open or closed. It is not accurate for different driver because of the individual differences on the size of eyes. We will discuss it in Experiments. In this paper, we take the individual differences into consideration to construct a unique classifier. It is equivalent to finding a soft threshold during the initialization process. It is not a fixed threshold, but an adaptive threshold. To make it works, we ask the driver to open his/her eyes and close his/her eyes for a while, obtain two sets of EAR via DCCNN and Dlib toolkit. A SVM classifier is trained with the input of two sets of data.

SVM [34] is a machine learning model that uses the structural risk minimization criterion. It is a linear classifier model with the largest interval defined in the feature space. Suppose given a training data set $S = \{(x_i, y_i),\ i = 1, 2, \ldots N\}$, where $x_i \in R^d$ is the input samples and $y_i \in \{+1, -1\}$ is the label corresponding to x_i. Assume x_i is a positive sample if $y_i = +1$ and on the contrary, it is a negative sample.

Generally, there is a linear discriminant function $f(x) = w^T x_i + b$ in the d-dimensional space to distinguish two types of data, and the classification hyperplane can be described as:

$$w^{*T} \cdot x + b^* = 0 \tag{7}$$

The normal vector w^T and the intercept b determine the discriminant function, so the hyperplane parameters w^T and b need to be calculated from the training data set. According to the basic idea of SVM, the constrained optimization problem of linear separable support vector machine can be obtained:

$$\begin{cases} \min\limits_{w,b} J(w) = \frac{1}{2}\,\|w\|_2^2 \\ s.t.\ y_i(w^T \cdot x_i + b) \geq 1,\ \ i = 1, 2, \ldots N \end{cases} \tag{8}$$

In the offline training module, for a specific driver, we collect two types of data when the eyes of driver are open and closed. Marking the eyes-open data as positive samples and eyes-closed data as negative samples. For a live video, the state of driver's eyes cannot change suddenly, that is, the eyes are almost impossible to change from open(closed) to closed(open) within one frame time. Therefore, we combine the EAR of six consecutive frames into one feature vector. According formula (8), we can get a unique classifier to judge whether the eyes are open or closed.

2) ONLINE MONITORING

Online monitoring module is the process of real-time detection during driving, which considering the individual differences of driver by application of the unique trained classifier. As shown in Figure 7, when the system starts up, a camera in front of the driver will capture the live video. All live video will be processed frame by frame. Firstly, the DCCNN is used to detect face of driver. If the region of face is obtained, it will be inputted to find the landmarks with Dlib toolkit. Otherwise, we assess the current frame is drowsy directly because the head of driver may not be forward-looking. Next, facial landmarks are obtained using Dlib toolkit. Similarly, if landmarks acquisition fails, the current frame is judged to a sleepy image. Finally, the EAR will be calculated according to the eyes landmarks. And the SVM classifier, which has trained in offline training module, is applied to decide whether the eyes of driver are open or closed with the input of EAR. The number of drowsy frames are stored with a parameter N_{drowsy}.

As mentioned, the driving drowsiness is a process of dynamic change. In [35], PERCLOS proved to be an effective indicator for driving drowsiness detection. The original PERCLOS method consists of two parts: (1) Judge whether the eyes are open or closed using P80, P70 or EM. Among of them, 1) P70, that is, if the eyelid coverage pupil area exceeds 70%, it is judged to be closed state; 2) P80, that is, the eyelid coverage pupil area exceeds 80%, then it is judged to be closed state; 3) EM, that is, if the eyelid coverage pupil area exceeds 50%, it is judged to be in the closed state. (2) Calculate the ratio of sleepy frames to total frames over time. In this paper, we have construct a unique classifier to

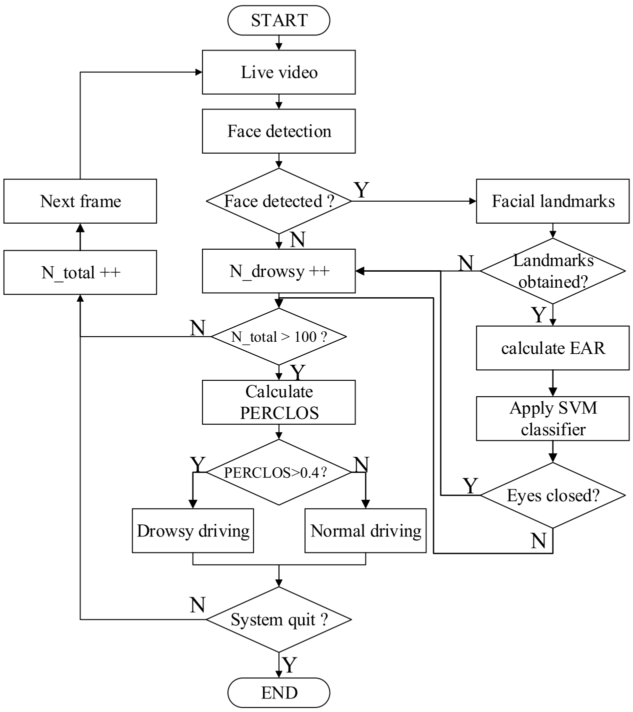


FIGURE 7. Flow chart of online monitoring.

judge the eyes state instead of using the hard threshold such as P80. Similar original PERCLOS part 2, we calculate the ratio of drowsy frames to total frames over time as well. PERCLOS can be computed by:

$$PERCLOS = \frac{N_{drowsy}}{N_{total}} \times 100\% \tag{9}$$

where N_{drowsy} is the number of drowsy frames judged by classifier and N_{total} is the total number of frames in a specific time.

In the online monitoring module, we set the N_{total} as 100 frames. If $PERCLOS > Th$, where Th is the threshold in a similar manner as [36], the driver is assessed driving drowsiness.

III. EXPERIMENTS

In this section, we first evaluate the effectiveness of the proposed DCCNN in Face Detection Data Set and Benchmark (FDDB) [37]. Then, we will discuss the correlation of EAR and the size of eyes to describe the individual differences of driver. At last, a series of comparative experiments are conducted to evaluate the performance of our proposed algorithm in both accuracy and speed.

A. ENVIRONMENT AND DATA SET

The experimental platform is Intel Core i5-7500 (main frequency: 3.4GHz) with x86 architecture, GTX1070TI (CUDA: 9.0; CUDNN: 7.0) with Pascal architecture, 16G DDR4 memory, opencv3.3.0 image library, deep learning computing framework is Tensorflow 1.7.

In this paper, we introduce two types of data sets to conduct experiments. The first one is FDDB

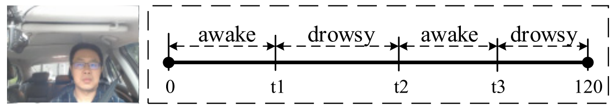


FIGURE 8. SDVD collected on simulator. Left: video clip of SDVD. Right: an annotated data. The driver is awake at 0 ∼ t1 and t2 ∼ t3. And he/she is drowsy at t1 ∼ t2 and t3 ∼ 120 (simulated states).

TABLE 1. Data sets.

Data set	Total images	Faces
FDDB	2845	5171
WIDER_FACE	30000	40000
ALFW	25000	25993

(http://vis-www.cs.umass.edu/fddb/index.html), which contains the annotations for 5171 faces in a set of 2845 images. The second is our own build video data sets, called Simulated Driving Video Data set (SDVD), which collected in simulated driving. For safety, we build the SDVD by asking the driver to make driving operations on the simulator, including the awake driving and the drowsy driving. SDVD contains 50 videos with different drivers, surroundings and times, etc. The subjects have different gender, age and eyes size. Each of them is 2 minutes and includes a manually marked awake period and sleepy period. As shown in Figure 8. Besides, in the process of DCCNN training, the WIDER_FACE dataset and AFLW dataset are applied to drive the training process. The information of the public datasets is shown in Table 1.

B. FACE DETECTION

1) QUALITATIVE DESCRIPTION

In order to evaluate the performance of the proposed face detection networks, we first conduct experiments in laboratory and real driving scenarios. The posture of head, light, background and facial decoration, etc. will have a great impact on face detection. Thus, we take all those factor into consideration to design the qualitative description experiment. As shown in Figure 9, the face can be detected accurately in laboratory (the first row of Figure 9), because of the ideal lighting and simple surroundings. In practical driving scenarios (the second and third rows of Figure 9), the DCCNN can obtain the face of driver as well, even the surroundings such as lighting is much more complicated than the laboratory.

2) QUANTITATIVE EVALUATION

In this section, we evaluate the effectiveness of the proposed DCCNN in FDDB. In machine learning, Intersection over Union (IOU) [38] and accuracy [39] are indicators for evaluating the performance of deep learning model.

IOU is a standard for measuring the accuracy of detecting corresponding objects in a specific data set. In object detection based on Image process, Ground-truth Bounding Box (GB) is the area labeled in data set, which represents the

FIGURE 9. Face detection in different scenarios.

true position of the object and Predicted Bounding Box (PB) is the region that predicted by deep learning model. IOU is computed by:

$$IOU = \frac{S(GB \bigcap PB)}{S(GB \bigcup PB)} \tag{10}$$

where $S(GB \bigcap PB)$ is the area of overlap and $S(GB \bigcup PB)$ is the area of union.

We introduce the IOU to indicate the similarity between the labeled position of face (Ground-truth Bounding Box) in FDDB data set and the predicted face position (Predicted Bounding Box). As shown in Figure 10. The higher the overlap of two bounding boxes, the higher of IOU value. Ideally, $IOU = 1$ when the predicted bounding box of face is same as the ground-truth bounding box in FDDB data set label. Generally speaking, it is considered that the object is detected if IOU above 0.5. In this paper, we assume that the face is detected correctly when $IOU > 0.7$.

To verify the performance of the face detection networks, we use the framework of multi-threaded input data provided by Tensorflow to combine the training data to disrupt the data sequence. In the training process, we set the batch size to 384, and the initial learning rate is 0.01. When the evaluation index is no longer improved, the learning rate is reduced by 10 times to 0.01, and the end rate is set to 0.00001. An epoch means that all training data are trained once. In this paper, we set the epoch to 50. After an epoch of training, a batch of test data from FDDB data set is fed to the network, to evaluate the performance of the model.

As mentioned, the DCCNN is a cascaded CNN of multi-task. The three sub-networks, Network-1, Network-2 and Network-3, can be trained and validated independently. The test accuracy curves of three sub-networks are shown in Figure 11, and the test loss curves are shown in Figure 12. With the increase of steps of training, the test accuracy is improved gradually. After 50 epochs, the final test accuracy of three sub-networks are 94.6%, 97.4% and 98.8%, because of the progressive architecture of DCCNN.

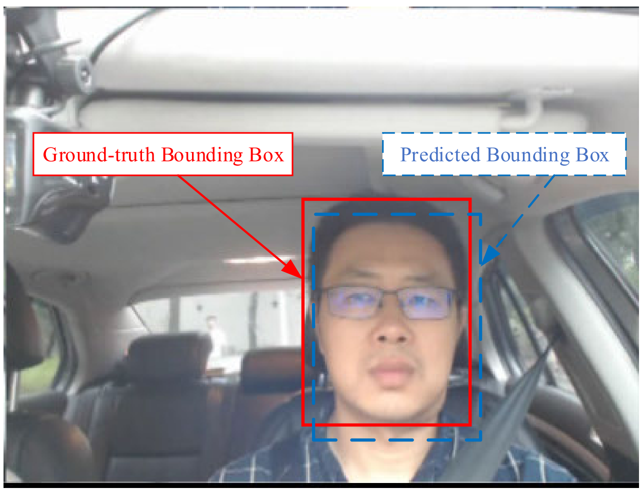


FIGURE 10. Demonstration of IOU in face detection.

In summary, the DCCNN we designed can capture the face of driver in various surroundings.

C. CORRELATION OF EAR AND THE SIZE OF EYES

Face detection and eyes landmarks location are the basis of driving drowsiness detection. A new parameter, EAR, is proposed in this paper to assess the state of driver's eyes. In order

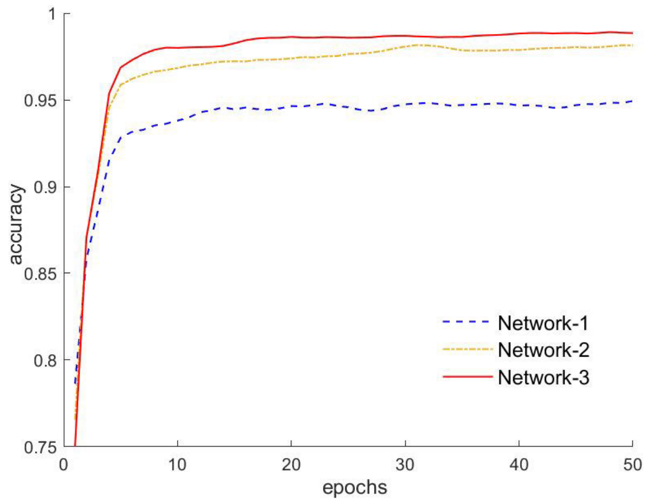


FIGURE 11. **Accuracy of the DCCNN.**

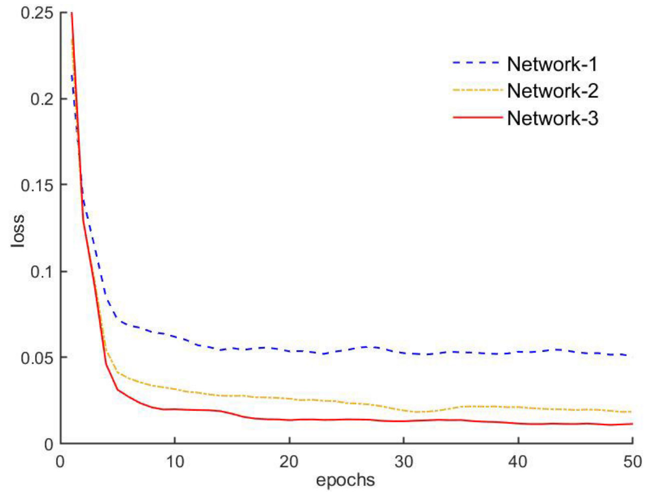


FIGURE 12. **Loss of the DCCNN.**

to verify the correlation of EAR and the size of eyes, which reflects the individual differences of driver, 4 subjects with different characteristics were selected. The size of eyes in case of natural open are measured in advance and size:1 ∼ size:4 are represent the 4 types of driver. Respectively, the size of eyes is 12mm, 16mm, 7mm and 20mm (size 1∼size 4). In the process of experiments, each subject was asked to open and close the eyes during the specific of time. The camera in front of driver capture the frames in real-time and the DCCNN is application to detect the face of driver. Then, calculate the EAR in the condition of opening and closure using Dlib toolkit.

Figure 13 shows the EAR curves of 4 subjects with different size of eyes. It can be seen from the figure that when the subject 1 (size: 1) is in the eyes opening state, the EAR is about 0.18. And the value is about 0.13 when the eyes of subject 1 are closed. At this time, the threshold of the driver eyes state classifier based on SVM is about 0.16 according to the optimal classifier principle. For subject 2 (size: 2), the EAR in different states are 0.3 and 0.15 respectively.

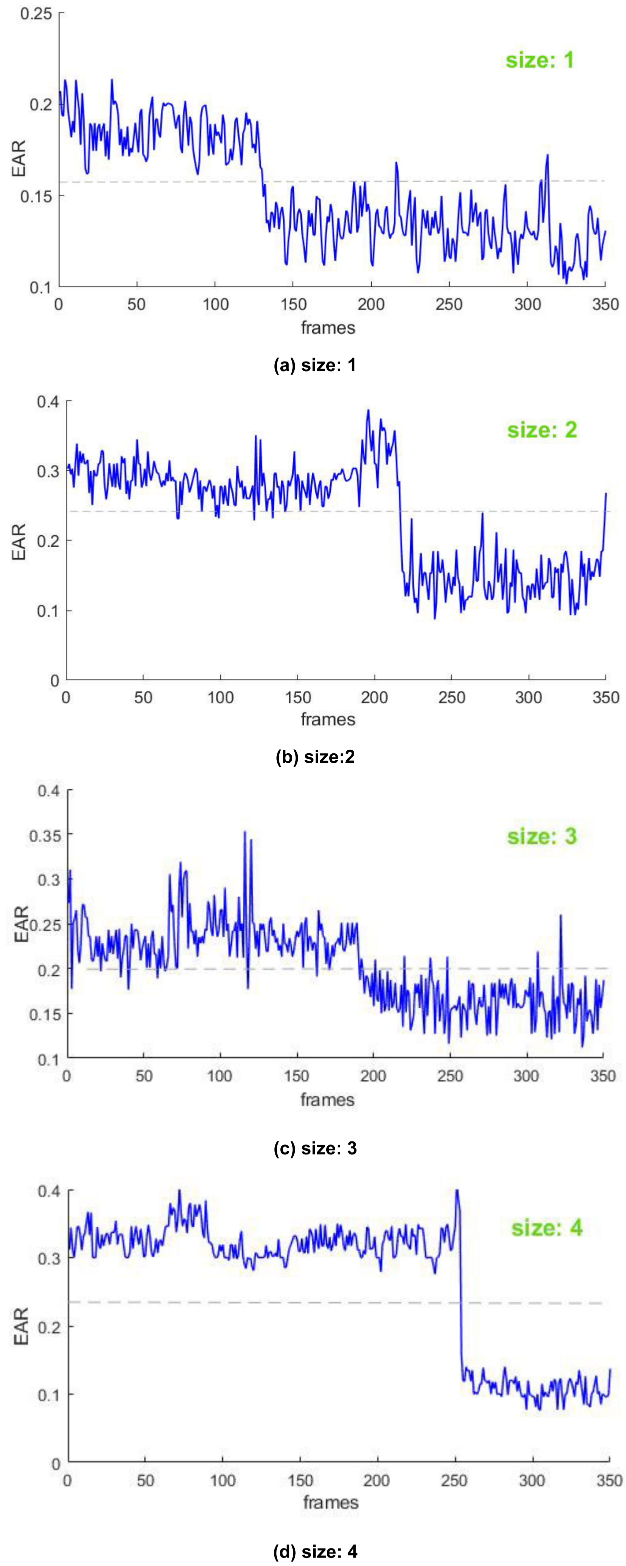


FIGURE 13. **EAR curve of 4 subjects.**

Therefore, the threshold of classifier is about 0.25. Particularly, the size of eyes between subject 3 and subject 4 are significant different to enhance our hypothesis. The size of object 3's eyes are small in naturally open condition so that the EAR of different states are approximate. In contrast, the EAR of object 4 in two types of state are significant different.

Therefore, the correlation of the EAR and the size of eyes is strong as we assumed above. Generally, the lager the driver's eyes are, the higher the EAR is. And the smaller the eyes are, the lower the EAR is. It is necessary to develop the driving drowsiness detection model with consideration of driver's individual differences, especially the size of eyes, to improve the performance of algorithm.

D. COMPARATIVE EXPERIMENTS

In order to verify the rationality of the driving drowsiness detection algorithm with consideration of individual differences, a series of comparative experiments are conducted. Firstly, we compared the accuracy on driving fatigue detection of using the hard threshold (0.8) and the method proposed in this paper. Then, we ask a driver to simulate the normal driving and drowsy driving state on the simulator, so that the change of eyes state and PERCLOS is visible clearly. Finally, we conduct another comparative experiment to compare the performance of our methods and relevant literatures both on accuracy and speed. For fair comparison, we use the same data and platform for both methods.

1) COMPARE WITH P80

We select 10 segments of simulated driving video randomly from the SDVD data set. The DCCNN is applied to detect the face of driver and then obtain the landmarks of face using the Dlib toolkit. On this basis, a comparative experiment was conducted.

Experiment 1: Calculate the EAR according to the landmarks of eyes. P80 is used to judge whether the eyes of driver are open or closed. That is, the eyes of driver are assessed to be closed if $EAR < 0.2$. Then, the driving state of driver is determined by the ratio of closed frames for a specific time (100 frames).

Experiment 2: Initializing the system firstly by training a unique SVM classifier with the input of the two sets of EAR value. Next, the driving state is assessed by using the online monitoring model proposed in this paper.

As mentioned above, for safety, we conduct the comparative experiments on the simulator. The SDVD data set contains the driving video with labeled information. For a sample, the driving state has been marked in advance as shown in Figure 8. The comparative experiments will detect the driving state and record the period of awake and drowsy. If the result obtained by the driving drowsiness model is consistent with the labeled information, it is considered that the model gets a correct evaluation of driving state. Therefore, the accuracy can be computed by:

$$accuracy = \frac{N_c}{N} \tag{11}$$

where N_c is the number of video clips correctly evaluated and N is the total number of experiment samples.

TABLE 2. Driver fatigue detection accuracy comparison.

accuracy/ %	P80	Ours
1	91.20	94.63
2	88.12	92.90
3	87.76	95.81
4	90.29	93.33
5	89.70	96.33
6	94.64	94.50
7	92.80	96.13
8	82.89	93.33
9	85.33	95.82
10	87.66	95.21
average	89.04	94.80

The comparative experiment is repeated 10 times. Table 2 shows the accuracy of driving drowsiness detection. It can be seen from the table that the algorithm we proposed improves the accuracy effectively.

2) STATE OF EYES AND PERCLOS IN DIFFERENT DRIVING CONDITION

We have description the correlation of EAR and the size of eyes with four types of driver. In order to visually evaluate the changes of driver's eyes state and the PERCLOS values under different driving conditions. We ask a driver (size: 1) to simulate two different driving states. As shown in Figure 14.

It can be seen from Figure.14 that when the driver is in normal driving state, the duration of the eyes opening is much longer than the time of closure. And the values of PERCLOS are all below the threshold (0.4). Conversely, the duration of closure is longer than the time of opening when the driver is in drowsy state. Most of the PERCLOS are over the threshold.

3) COMPARE WITH RELEVANT LITERATURES

Figure 15 shows the calculation speed of each module of the algorithm during the experiment of the four subjects. The experimental video stream is 640∗480 resolution and the frame rate is 30fps. It can be seen from the figure that the average speed of the face detection module is 35.95ms/f, and the average speed of the landmarks obtaining module is 13.80 ms/f. The average speed of the fatigue driving detection algorithm we proposed is 49.75 ms/f.

In order to furtherly verify the performance of the proposed algorithm, we use the self-built data set, SDVD, to compare our algorithm with the drowsy detection algorithms [40] [41]

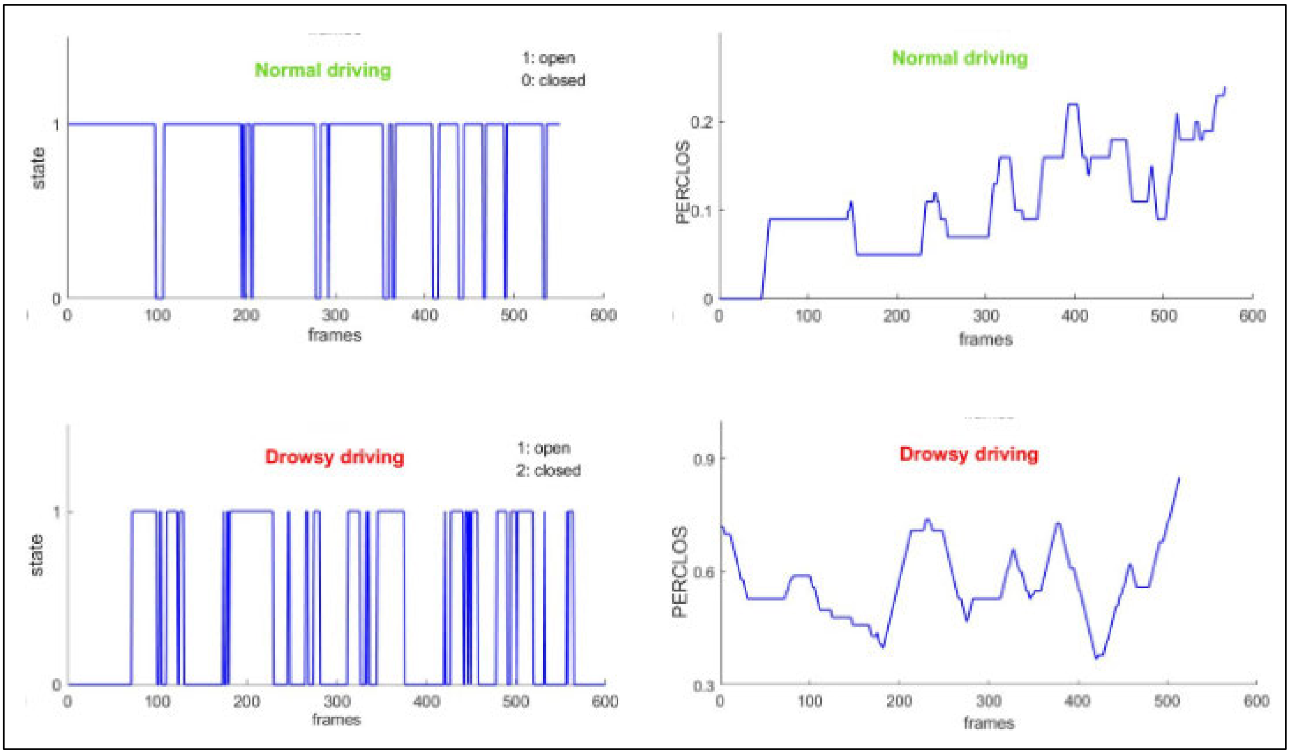


FIGURE 14. Eyes state and PERCLOS under different driving conditions.

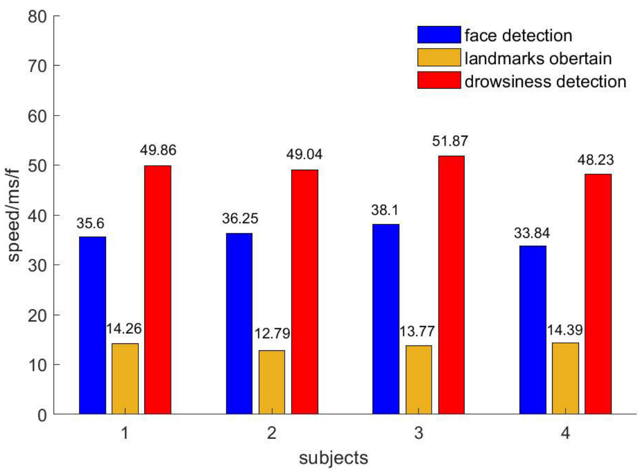


FIGURE 15. Algorithm speed.

TABLE 3. Comparison.

Algorithms	Accuracy/ %	speed/ ms·f -1
Adaboost	92.10	58.61
LSTM	91.48	65.64
Ours	94.80	49.75

proposed in recent years. The detection accuracies and speeds are shown in Table 3.

IV. CONCLUSION

Research on drowsy driving detection algorithm is one of the most important methods to reduce traffic accidents. As we know, there are significant individual differences, especially the size of eyes, between different person. It is crucial to take the individual differences into consideration when study on the algorithm based on computer vision.

In this paper, we propose a new driving drowsiness detection algorithm with consideration of individual differences. Firstly, we design a deep cascaded convolutional neural network model named DCCNN, which avoids the process of artificial feature extraction in traditional face detection algorithms, to obtain the face of a driver in live video. The performance of the model is tested by qualitative description and quantitative evaluation. Experimental results show that the accuracy of face detection can reach at 98.8%. Secondly, we propose a new parameter, EAR, based on the Dlib toolkit, to assess the state of driver's eyes. Compared with the traditional methods, the EAR is more stable thanks to the Cascaded Pose Regression algorithm. Experimental results show that there is a very strong correlation between the EAR and the size of a driver's eyes, which proves the rationality of our ideas. Finally, taking the individual differences of the drivers into consideration, we construct the offline training module and online monitoring module in the paper. A unique classifier based on SVM is trained for a specific driver and the state of eyes is judged with the application of the pre-trained classifier during driving. Experimental results demonstrate that our methods consistently outperform the state-of-the-art methods on the public data set and self-built data set while keeping real-time performance.

In the future, we will focus on the following topic: 1) explore the multi-feature such as mouth and head posture fusion methods to further improve the algorithm performance. 2) conduct the driving drowsiness detection research at nighttime because it is easier to drowsy driving at night.

ACKNOWLEDGMENT

This work was partially support by "Guangdong Natural Science Foundation 2020- Front Vehicles Perception and Cooperative Control of Safety Distance for Driverless Vehicle at Nighttime".

REFERENCES

[1] A. Amodio, M. Ermidoro, D. Maggi, S. Formentin, and S. M. Savaresi, "Automatic detection of driver impairment based on pupillary light reflex," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 8, pp. 3038–3048, Aug. 2019, doi: 10.1109/TITS.2018.2871262.

[2] X. Li, X. Lian, and F. Liu, "Rear-end road crash characteristics analysis based on Chinese in-depth crash study data," in *Proc. Cota Int. Conf. Transp. Prof.*, 2016, pp. 1536–1545.

[3] G. Zhang, K. K. Yau, X. Zhang, and Y. Li, "Traffic accidents involving fatigue driving and their extent of casualties," *Accident Anal., Prevent.*, vol. 87, pp. 34–42, Feb. 2016, doi: 10.1016/j.aap.2015.10.033.

[4] D. Mollicone, K. Kan, C. Mott, R. Bartels, S. Bruneau, M. van Wollen, A. R. Sparrow, and H. P. A. Van Dongen, "Predicting performance and safety based on driver fatigue," *Accident Anal., Prevention*, vol. 126, pp. 142–145, May 2019, doi: 10.1016/j.aap.2018.03.004.

[5] X. Sun, H. Zhang, W. Meng, R. Zhang, K. Li, and T. Peng, "Primary resonance analysis and vibration suppression for the harmonically excited nonlinear suspension system using a pair of symmetric viscoelastic buffers," *Nonlinear Dyn.*, vol.94, no. 2, pp. 1243–1265, 2018, doi: 10.1007/s11071-018-4421-9.

[6] Z. Ning, P. Dong, X. Wang, M. S. Obaidat, X. Hu, L. Guo, Y. Guo, J. Huang, B. Hu, and Y. Li, "When deep reinforcement learning meets 5G vehicular networks: A distributed offloading framework for traffic big data," *IEEE Trans. Ind. Informat.*, to be published, doi: 10.1109/TII.2019.2937079.

[7] F. You, Y.-H. Li, L. Huang, K. Chen, R.-H. Zhang, and J.-M. Xu, "Monitoring drivers' sleepy status at night based on machine vision," *Multimed. Tools Appl.*, vol. 76, no. 13, pp. 14869–14886, 2017, doi: 10.1007/s11042-016-4103-x.

[8] G. Sikander and S. Anwar, "Driver fatigue detection systems: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2339–2352, Jun. 2019, doi: 10.1109/TITS.2018.2868499.

[9] X. Li, L. Hong, J. Wang, and X. Liu, "Fatigue driving detection model based on multi-feature fusion and semi-supervised active learning," *IET Intell. Transp. Syst.*, vol. 13, no. 9, pp. 1401–1409, Sep. 2019.

[10] G. Niu and C. Wang, "Driver fatigue features extraction," *Math. Problems Eng.*, vol. 2014, pp. 1–10, Jun. 2014.

[11] Z. Ning, Y. Feng, M. Collotta, X. Kong, X. Wang, L. Guo, X. Hu, and B. Hu, "Deep learning in edge of vehicles: Exploring trirelationship for data transmission," *IEEE Trans. Ind. Informat.*, vol. 15, no. 10, pp. 5737–5746, Oct. 2019.

[12] M. Khan and S. Lee, "A comprehensive survey of driving monitoring and assistance systems," *Sensors*, vol. 19, no. 11, p. 2574, 2019.

[13] A. S. Zandi, A. Quddus, L. Prest, and F. J. E. Comeau, "Non-intrusive detection of drowsy driving based on eye tracking data," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2673, no. 6, pp. 247–257, May 2019.

[14] L. M. Bergasa, J. Nuevo, M. A. Sotelo, R. Barea, and M. E. Lopez, "Real-time system for monitoring driver vigilance," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 1, pp. 63–77, Mar. 2006.

[15] X. Zhao, "Eye feature point detection based on single convolutional neural network," *IET Comput. Vis.*, vol. 12, no. 4, pp. 453–457, 2018.

[16] F. Zhang, J. Su, L. Geng, and Z. Xiao, "Driver fatigue detection based on eye state recognition," in *Proc. Int. Conf. Mach. Vis. Inf. Technol.*, Feb. 2017, pp. 105–110, doi: 10.1109/CMVIT.2017.25.

[17] Z. Xiao, Z. Hu, L. Geng, F. Zhang, J. Wu, and Y. Li, "Fatigue driving recognition network: Fatigue driving recognition via convolutional neural network and long short-term memory units," *IET Intell. Transp. Syst.*, vol. 13, no. 9, pp. 1410–1416, Sep. 2019.

[18] W. Liu, "Convolutional two-stream network using multi-facial feature fusion for driver fatigue detection," *Future Internet*, vol. 11, no. 5, p. 115, 2019.

[19] Z. Ning, P. Dong, X. Wang, J. Rodrigues, and F. Xia, "Deep reinforcement learning for vehicular edge computing: An intelligent offloading system," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 6, 2019, Art. no. 60.

[20] X. Wang and C. Xu, "Driver drowsiness detection based on non-intrusive metrics considering individual specifics," *Accident Anal. Prevention*, vol. 95, pp. 350–357, Oct. 2016.

[21] L. Zhang, F. Liu, and J. Tang, "Real-time system for driver fatigue detection by RGB-D Camera," *ACM Trans. Intell. Syst. Technol. (TIST)*, vol. 6, no. 2, pp. 1–17, 2015.

[22] R. Fu, H. Wang, and W. Zhao, "Dynamic driver fatigue detection using hidden Markov model in real driving condition," *Expert Syst. Appl.*, vol. 63, pp. 397–411, Nov. 2016.

[23] Y. Jiang, S. Guo, and S. Deng, "Denoising and chaotic feature extraction of electrocardial signals for driver fatigue detection by kolmogorov entropy," *J. Dyn. Syst., Meas. Control, Trans. ASME*, vol. 141, no. 2, 2019, Art. no. 0210131, doi: 10.1115/1.4041355.

[24] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499–1503, Oct. 2016.

[25] Z. Ning, J. Huang, X. Wang, J. Rodrigues, and L. Guo, "Mobile edge computing-enabled Internet of Vehicles: Toward energy-efficient scheduling," *IEEE Netw.*, vol. 33, no. 5, pp. 198–205, Sep./Oct. 2019.

[26] H. Xiong, X. Zhu, and R. Zhang, "Energy recovery strategy numerical simulation for dual axle drive pure electric vehicle based on motor loss model and big data calculation," *Complexity*, vol. 2018, Aug. 2018, Art. no. 4071743, doi: 10.1155/2018/4071743.

[27] S. Yang, P. Luo, C. C. Loy, and X. Tang, "WIDER FACE: A face detection benchmark," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 5525–5533, doi: 10.1109/CVPR.2016.596.

[28] M. Köstinger, P. Wohlhart, P. M. Roth, H. Bischof, "Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Nov. 2011, pp. 2144–2151, doi: 10.1109/ICCVW.2011.6130513.

[29] Z. Ning, J. Huang, and X. Wang, "Vehicular fog computing: Enabling real-time traffic management for smart cities," *IEEE Wireless Commun.*, vol. 26, no. 1, pp. 87–93, Feb. 2019.

[30] P. Dollár, P. Welinder, and P. Perona, "Cascaded pose regression," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 1078–1085, doi: 10.1109/CVPR.2010.5540094.

[31] D. E. King, "Dlib-ml: A machine learning toolkit," *J. Mach. Learn. Res.*, vol. 10, pp. 1755–1758, Jan. 2009.

[32] A. Tolba, "Content accessibility preference approach for improving service optimality in Internet of Vehicles," *Comput. Netw.*, vol. 152, pp. 78–86, Apr. 2019.

[33] X. Kong, F. Xia, J. Li, M. Hou, M. Li, and Y. Xiang, "A shared bus profiling scheme for smart cities based on heterogeneous mobile crowdsourced data," *IEEE Trans. Ind. Informat.*, to be published, doi: 10.1109/TII.2019.2947063.

[34] J. Xu, W. Zeng, Y. Lan, J. Guo, and X. Cheng, "Modeling the parameter interactions in ranking SVM with low-rank approximation," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 6, pp. 1181–1193, Jun. 2019.

[35] A. Tolba, "Trust-based distributed authentication method for collision attack avoidance in VANETs," *IEEE Access*, vol. 6, pp. 62747–62755, 2018.

[36] R.-H. Zhang, Z.-C. He, H.-W. Wang, F. You, and K.-N. Li, "Study on self-tuning tyre friction control for developing main-servo loop integrated chassis control system," *IEEE Access*, vol. 5, pp. 6649–6660, 2017.

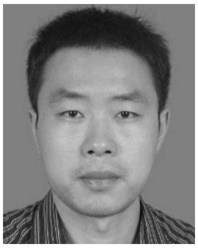
[37] X. Sun, P. Wu, and S. C. H. Hoi, "Face detection using deep learning: An improved faster RCNN approach," *Neurocomputing*, vol. 299, pp. 42–50, Jul. 2018.

[38] X. Kong, X. Liu, B. Jedari, M. Li, L. Wan, and F. Xia, "Mobile crowdsourcing in smart cities: Technologies, applications, and future challenges," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 8095–8113, Oct. 2019.

[39] A. Tolba, O. Said, and Z. Al-Makhadmeh, "MDS: Multi-level decision system for patient behavior analysis based on wearable device information," *Comput. Commun.*, vol. 147, pp. 180–187, Nov. 2019.

[40] L. Geng, "Real-time driver fatigue detection based on morphology infrared features and deep learning," *Hongwai Yu Jiguang Gongcheng/Infrared Laser Eng.*, vol. 47, no. 2, 2018, Art. no. 203009.

[41] J. Guo and H. Markoni, "Driver drowsiness detection using hybrid convolutional neural network and long short-term memory," *Multimedia Tools Appl.*, vol. 78, no. 20, pp. 29059–29087, 2019.



FENG YOU was born in Zunyi, Guizhou, in 1977. He received the B.S. degree in automobile transportation engineering from Guizhou University, Guizhou, in 1995, and the M.S. and Ph.D. degrees in automobile transportation engineering from Jilin University, Changchun, in 2005. From 2005 to 2011, he was a Lecturer with the School of Civil Engineering and Transportation, South China University of Technology. Since 2012, he has been an Associate Professor with the School of Civil Engineering and Transportation, South China University of Technology. He is the author of more than 50 articles, and more than ten inventions. His research interests include connected vehicle, advanced driver assistance, traffic information and safety, cooperative safety control, and vehicle active safety and driver behavior. He is a Reviewer of the Journal of *Safety Science*, *Accident Analysis & Prevention*, and *Jilin University Engineering and Technology Edition*.

XIAOLONG LI was born in Bengbu, Anhui, in 1993. He received the B.E degree from Southeast University, Nanjing, China, in 2016. He is currently pursuing the master's degree in communication and transportation engineering with the South China University of Technology, Guangzhou, China. His interests include intelligent vehicles, computer vision, and driving behavior analysis.

YUNBO GONG was born in Tieling, Liaoning, in 1995. He received the B.E. degree from the South China University of Technology, Guangzhou, China, in 2018, where he is currently pursuing the master's degree in traffic information engineering and control. His interests include intelligent vehicles, computer vision, and 3D laser radar.

HAIWEI WANG received the B.E. and M.S. degrees from Jilin University, Changchun, China, and the Ph.D. degree from the South China University of Technology. She is currently a Lecturer with the School of Transportation and Economic Management, Guangdong Communication Polytechnic, Guangzhou, China. She has published four articles in international journals. Her current research interests include ITS and vehicle control.

HONGYI LI was born in Urumqi, Xinjiang, China, in 1986. He received the B.S. degree in mechanic from the Beijing Institute of Technology, Beijing, in 2008, and the Ph.D. degree in physics and chemistry materials from the Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Urumqi, in 2013. From 2011 to 2012, he was a dual culture Ph.D. with Northwestern University, Chicago, USA. From 2013 to 2014, he was a Senior Engineer. Since 2015, he has been a Professor with the Xinjiang Quality of Products Supervision and Inspection Institute of Technology. Besides, he was a Visit Research Fellow with the Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences. His research interests include materials science and information technology.