IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received August 18, 2020, accepted September 7, 2020, date of publication September 22, 2020, date of current version October 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3025818

Driver Fatigue Detection Method Based on Eye States With Pupil and Iris Segmentation

QIANYANG ZHUANG, ZHANG KEHUA, JIAYI WANG, AND QIANQIAN CHEN
Key Laboratory of Urban Rail Transit Intelligent Operation and Maintenance Technology and Equipment of Zhejiang Province, Zhejiang Normal University, Jinhua 321005, China

Corresponding author: Zhang Kehua (mature@zjnu.edu.cn)

This work was supported in part by the Natural Science Foundation of Zhejiang Provincial of China under Grant LY17F050002, and in part by the National Natural Science Foundation of China under Grant 61205012.

ABSTRACT Fatigue driving has become one of the most common causes for traffic accidents. In this article, we proposed an effective fatigue detection method based on eye status with pupil and iris segmentation. The segmented feature map can guide the detection to focus on pupil and iris. A streamlined network, consisting of a segmentation network and a decision network, is designed, which greatly improves the accuracy and generalization of eye openness estimation. Specifically, the segmentation network that uses light U-Net structure performs a pixel-level classification on the eye images, which can accurately extract pupil and iris features from the video's images. Then, the extracted feature map is used to guide the decision network to estimate eye openness. Finally, the detection method is test by the National Tsing Hua University Drowsy Driver Detection (NTHU-DDD) Video Dataset and the precision of fatigue detection achieves 96.72%. Experimental results demonstrate that the proposed method can accurately detect the driver fatigue in-time and possesses superior accuracy over the state-of-the-art techniques.

INDEX TERMS Fatigue detection, pupil and iris segmentation, percentage of eyelid closure (PERCLOS), deep learning.

I. INTRODUCTION

Driving for a long time will cause driver fatigue. And the driver will have a phenomenon of drowsiness, which easily leads to a traffic accident. In the 2018 Global Road Safety Status Report, World Health Organization (WHO) noted that 1.35 million people died in road traffic accidents every year. According to the different fatigue performances, there are the following three kinds of detection methods: (1) fatigue detection method based on the physiological signals, which is to judge fatigue according to the oxygen level of brain tissues and muscle tissues [1], [2], (2) fatigue detection method based on the vehicle behavior, which monitors the variations of steering wheel angle, lane position, speed, acceleration, and braking to predict the driver fatigue [3], [4], and (3) fatigue detection method based on the drivers' facial features, which monitors the facial expressions to judge the driver's fatigue state because the frequency of yawning and blinking increases when the driver is fatigued [5], [6]. However, the first two kinds of methods, which are intrusive detection methods, will affect the driving behavior. And the third kind of methods, which only depend on vision analysis to monitor the driver's status, is non-intrusive and does not interfere with the driving behavior. Besides, Dinge and Grace [7] have found that PERCLOS (percentage of eyelid closure over the pupil) was an effective indicator for fatigue detection.

In order to accurately calculate PERCLOS, we must judge the eye status in real time. Therefore, a large number of researchers have explored methods of eye status detection [8]–[10]. As shown in Fig. 1, in most of them the eye images are directly taken as the input, and then the Convolutional Neural Networks (CNN) is used to judge to the eye status. To ensure strong generalization ability of the network, the deep learning network models has also been established. However, in order to achieve the real-time detection, the deep learning network level should not be too deep [11]. In response to this trade-off, we proposed a segmentation-based eye status detection model (SESDM). The model consists of a segmentation network and a decision network. It not only improves the accuracy of the network, but also improves the generalization of the network.

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

FIGURE 1. General framework for eye status detection.

A. RELATED WORK

Accurate eye status detection method is very important for fatigue detection. In this part, we will introduce the currently popular eye status detection methods, which are based on shape recognition, template matching, traditional machine learning and deep learning, respectively.

Shape-based approaches aim to recognize the eye states according to the geometric relationships or circular shape of visible iris. Kurylyak *et al.* [12] set some thresholds for the difference between video frames in eye region pixel level to detect the eyelid movement. Then the differences of vertical and horizontal projections are used to detect the degree of eye openness. In addition, Wang *et al.* [13] proposed the method of blink detection based on the trained adaptive boosting (Adaboost) with contour circle (CC). They experimentally proved that the vertical coordinate of the CC can be the best feature which can classify open/closed eyes by the linear decision surface.

Template matching is another simple and direct method of eye status detection. Yutian *et al.* [14] used template matching to roughly detect the driver's eye status and then the curvature of the upper eyelid was captured for finer identification. Gonzalezortega *et al.* [15] used projection operations to generate three templates for open, nearly closed, and closed eyes, respectively. And the state of the eye can be judged by comparing the similarity between the templates and the test image.

However, using shape recognition or template matching, the different parameters need to be set for different individuals. Traditional machine learning method can effectively overcome this shortcoming. Eddine *et al.* [16] used the support vector machine (SVM) and multi-layer perceptron (MLP) to identify local binary pattern (LBP) feature images. The model is verified on the ZJU dataset with an accuracy of 95%. Han *et al.* [17] developed a driver sleepiness detection system. This system predicts the state of the eye through a single-layer neural network regression model. They collected a 30-minute driving video of 8 people, with an accuracy of 91%.

In a real environment, varying lighting conditions will lead to increased difficulty in eye status detection. Template matching, shape recognition, and traditional learning method have poor detection performance in the complex environment. With the development of deep learning, computer vision has achieved breakthrough results [18]–[20]. In the 2012 ImageNet competition, Krizhevsky *et al.* [20] achieved the outstanding results in building the AlexNet, because CNN has powerful feature extraction capabilities. The method of using deep learning to recognize the state of the eye has been proposed. Wang *et al.* [11] established a dual-stream bidirectional CNN model for eye status detection. The network uses binocular consistency as the information exchange. Finally, using the interactive information, the state of the eyes can be judged. However, using the projection features of pupil to judge the eye status needs to adjust the parameters according to differences among people, which leads to poor generalization ability. To overcome the influence of wearing glasses, Zhang *et al.* [10] captured the driver's facial videos by near-infrared cameras and then used the CNN to judge the status of eyes.

In summary, shape recognition and the template matching methods have the problem of insufficient generalization. Thus, it is very difficult to set a common threshold and template. Traditional machine learning methods can divide the eyes into open, closed, and half-open states. However, it is fuzzy for the method to judge the process from closed to open. Hence, to accurately characterize the driver drowsiness based on the eye images, it is desirable to measure the eye-openness continuously. We developed a novel method for eye-openness estimation based on the pupil and iris of the eye so that there was no need to detect the feature points and curves around the eyes. The SESDM, consisting of a segmentation network and a decision network, was designed. Specifically, the segmentation network performs pixel-level classification of eye images, which can accurately extract pupil and iris features from the video's images. Finally, this prior feature is used to guide the decision network to estimate the eye openness. Accurate and timely driving fatigue detection can be achieved.

II. METHOD

In order to reduce the interference of light changes, we used infrared cameras to capture facial videos. The proposed fatigue detection method is as shown in Fig. 2, which is divided into three stages. In the first stage, the Dlib face key point detection algorithm was adopted, which can detect 68 key points of the face, including the edge position information of the eyes, nose and mouth. The eye image twice the width of the eye can be selected and cropped. By adaptive cropping of the eye images, the robustness of the fatigue method is greatly enhanced. In the second stage, the degree of the eye openness is successfully estimated by the proposed SESDM network. As shown in Fig. 3, this part includes two networks. The segmentation network, as the first network, is used to segment the original eye image at the pixel level. The pupil and iris features in the eye image are extracted. Then the prior features and the original images are input into the second network, called the decision network, which determine the level of eye openness. In the third stage, SESDM is used to estimate the eye openness of each frame. And we calculate the PERCLOS value based on the level of eye openness.

A. SEGMENTATION NETWORK

Pupil and iris features play a key role in the task of eye status detection. In the recent method [11], the author introduces the

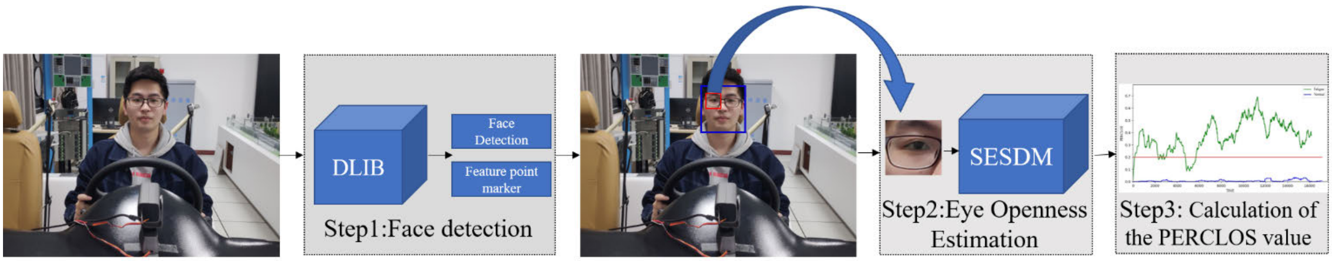


FIGURE 2. Fatigue detection framework.

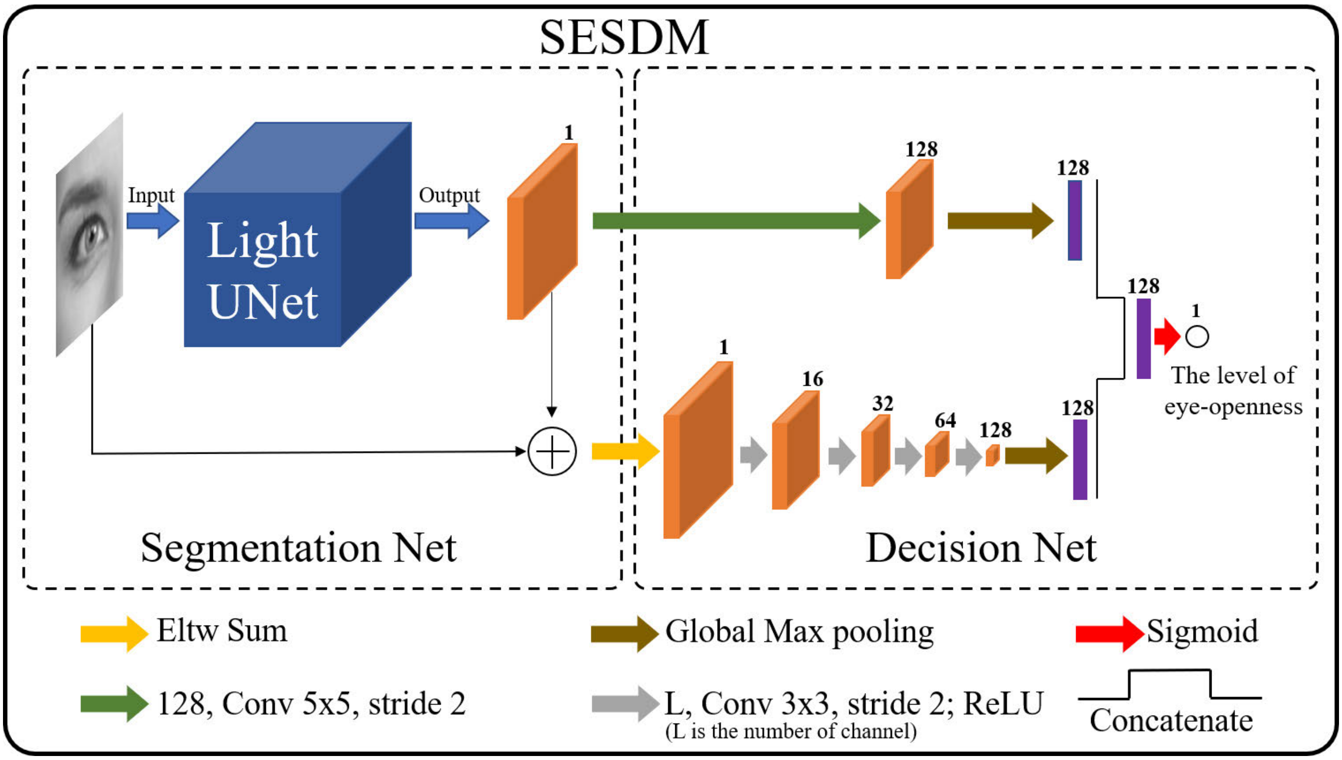


FIGURE 3. Eye state estimation network structure.

pupil feature in the network. But the method of pupil projection is susceptible to be affected by the environment. Thus, here, a segmentation network as a preprocessing method was proposed. As shown in Fig. 4, the U-Net for the pupil and iris segmentation were adopted, which is a kind of the Fully Convolutional Networks (FCN). It is suitable for our work because it can perform the classification and localization in real-time [21]. The segmentation network converts pupil and iris to the foreground, and the other areas to the background. The shortcut of the segmentation output map is introduced to fully express the characteristics of the level of eye openness. As shown in Fig. 4, the left side of the U-Net structure is called contracting path, which consists of 3×3 convolutional layers and max pooling layers, and the right side is called expansive path, which consists of convolutional layers and concatenating paths from the layers from the contracting path. However, in the case of U-Net, all points around the pupil and iris are classified as a region of interest. The left side enables the multiscale analysis by stacking the image pyramids, and the right side enables to acquire the pixel-wise classification results. The U-Net also concatenates the reduced feature map in the contracting path and upsampled output that enables the output to assemble the information of various scales. This resolves the trade-off between localization accuracy and the abundance of context mentioned above. Moreover, since there are no fully connected layers, the network requires a smaller number of parameters compared to the regression network at the same depth.

B. DECISION NETWORK

The segmentation network's input and single-channel output are summarized point to point with equivalent weights.

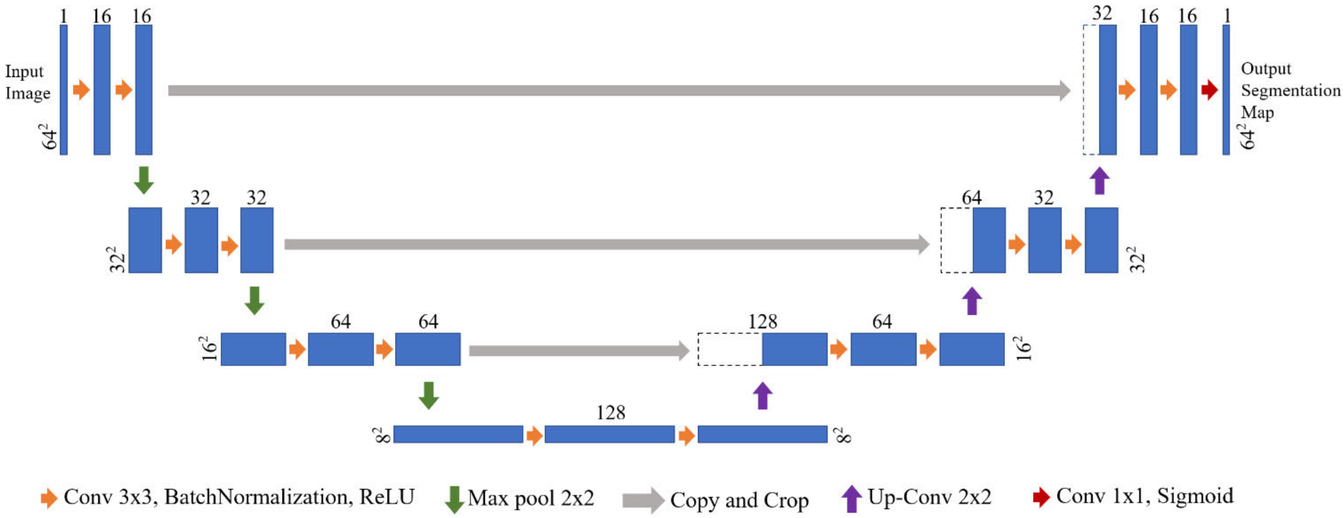


FIGURE 4. A light U-Net architecture for pupil and iris segmentation. Each rectangular box represents a feature map, and the number on the box represents the number of channels in the feature, the number before the box represents the size of the feature.

This results in a new layer that represents the input for the remaining layers with a convolutional layer with 3×3 kernel sizes and a ReLU layer. Combination of both layers is repeated 4 times, with 16, 32, 64 and 128 channels in the first, second, third and fourth convolutional layers, respectively. A detailed depiction of the architecture is given in Fig. 3. The number of channels is increased as the resolution of the features decreases, therefore resulting in the same computational requirement for each layer. The proposed design effectively results in a 16-times-smaller resolution of the last convolutional layer than that of the original image. Next, the network performs global maximum, resulting in 128 output neurons. Then, the segmentation output map performs convolution and global maximum operations, resulting in 128 output neurons. Finally, we connect two 128 neurons and map them into a new 128 output neurons. This design results in 128 output neurons that are combined with the weights into the final output neurons.

The design of the decision network follows two important principles. First, the appropriate capacity for complex shapes is ensured by using several layers of convolution. This enables the network to capture not only the local shapes, but also the global ones that span a large area of the image. Second, the decision network uses not only the segmentation network's output map, but also the new feature map which is generated by the input and output of segmentation network. This introduces a shortcut that the network can avoid using a large number of feature maps. It can also guide the network to focus on pupil and iris. The shortcuts are implemented at two levels. One is at the beginning of the decision network where the feature map is fed into several convolutional layers of the decision network, and another one is at the end of the decision network where the segmentation output map performs convolution and global maximum operations, resulting in 128 output neurons.

C. FATIGUE PARAMETER

PERCLOS is still considered as the most effective measurement parameter of drivers fatigue for vision-based non-intrusive approaches [22]–[25]. PERCLOS is defined as the percentage of eyelid closure over the pupil over a specified time period. Specifically, PERCLOS calculates the proportion of time within a specified time duration that the eyelid covers over 80% of the pupil [25]. As shown in Fig. 5, it is a schematic diagram of eye status. As shown in Formula (1), we derive the eye aspect ratio (EAR) that is used as an estimate of the eye openness.

$$EAR = \frac{h}{w} \tag{1}$$

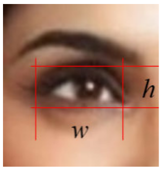


FIGURE 5. The diagram of eye. ***w*** represents the width of the eye, and ***h*** represents the distance between the eyelids.

According to Ref. [26], for a fully opened eye, the eye openness level EAR is around 0.5 and 80% of the eyelid closure corresponds to EAR = 0.1. Therefore, the PERCLOS defined on eye openness can be computed as:

$$F_{PERCLOS} = \frac{n_{close}}{N_{total}} * 100\% \tag{2}$$

where n_{close} represents the number of frames with closed eyes, and N_{total} represents the total number of detected frames in the specified time. Therefore, according to the different PERCLOS values [27], the status can be divided into normal and fatigue, the formula is:

$$S_t = \begin{cases} Fatigue & if\ F_{perclos} \geq 0.2 \\ Normal & if\ F_{perclos} < 0.2 \end{cases} \quad (3)$$

where 0.2 is the PERCLOS threshold for dividing normal and fatigue status.

III. DATASET

Everyone has different eye characteristics. Therefore, we need large amount of tag data for training an excellent eye-openness estimation network. Three facial key point datasets, include WFLW [28], the 300-W Challenge [29]–[31] and Helen [32], are used to obtain eye data. In addition to the key points, there are also information such as different ethnicities, lighting conditions and whether wearing glasses. Among them, the WFLW dataset is a dataset with 98 key points, and both the 300-W Challenge and HELEN dataset mark 68 key points. The level of eye openness information and the foreground can be obtained by its mark file. Maps of the pupil and iris, as shown in Fig. 6(b), are used to calculate the level of eye openness through formula (4).

$$\mathrm{EAR} = \begin{cases} \dfrac{\|p_3 - p_7\|}{\|p_1 - p_5\|} & if\ Landmark = 98 \\ \dfrac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\,\|p_1 - p_4\|} & if\ Landmark = 68 \end{cases} \quad (4)$$

where $p_1, \ldots, p_6$ are the 2D landmark locations, depicted in Fig. 6(b).

In order to ensure the data quality, we select the eye image whose resolution is greater than 45 x45. Finally, we collected 9728 eye data, including 6225 data as the training set, 1557 data as the validation set, and 1946 data as the test set, named as "EYE DATA".

IV. IMPLEMENTATION

Our training environment is on a server computer with NVIDIA TITAN XP GPU. When training the segmentation network, we convert the images' resolution to 64 × 64. Additionally, the input images are normalized through mean channel subtraction. Optimization is performed using Adam with a mini-batch size of 64 and a momentum of 0.9. The initial learning rate is set to 0.001. The network is trained for 100 epochs on the training set. This process is binary classification for each pixel of the original image. Finally, the eye image is divided into foreground and background, where the foreground contains pupil and iris.

The decision network is used to estimate the eye openness. The training is carried out separately from the segmentation network. Specifically, first of all, only the segmentation network for independent training to optimize the weight parameters. Then, the decision network is trained, which can avoid the over-fitting problem caused by a large number of weights in the segmentation network. And the segmentation network provides prior knowledge to the decision network, which makes the decision network more effective in feature extraction. When training the decision network, the same data processing method was used as in the segmentation network. Optimization is performed by using Adam with a mini-batch size of 64. The initial learning rate is set to 0.0001. The network is trained for 100 epochs on the training set. We save the best model by evaluating on the validation set during the optimization process. At test time, the model predicts the level of eye openness of each eye image.

FIGURE 6. The production process of EYE DATA: (a) Extract the eye image. (b) Calculate the level of eye openness according to formula (4). (c)annotating binary ground-truth map.

V. EXPERIMENTS

In this section, we analyze the influence of different structural parameters on the performance of SESDM, compare

TABLE 1. Test set identification results of SESDM with different U-Net structures in EYE DATA.

MODELS	The number of floors of U-Net	The channels on the 1st floor of U-Net	The number of samples	Mean Error	Error<=0.1	Error>0.1
SESDM 3_16	3	16	1946	0.039	1839	107
SESDM 3_32	3	32	1946	0.037	1860	86
SESDM 4_16	4	16	1946	0.036	1863	83
SESDM 4_32	4	32	1946	0.042	1833	113

the performance of SESDM network with other typical networks for eye openness estimation, and compare the performance of our method with other fatigue detection methods. Finally, we test the performance of our method in real driving environments.

A. COMPARING THE PERFORMANCE OF SESDM WITH DIFFERENT PARAMETERS

In this subsection, we analyze the performance vs. network complexity. As stated previously, the segmentation network uses a light U-Net architecture, compared to the original U-Net network. However, the structures with different complexity will affect the model performance. We use the "EYE DATA" dataset to train models with different structures and compare their performances. As shown in Table 1, our U-Net with the number of floors and/or the number of channels, are compared. Fig. 7 shows the change curve of the recognition accuracy of the U-Net and the loss of SESDM on the validation set. It can be seen that when U-Net's number of floors is 4 and channels on the 1st floor is 16, it performs best on the validation set, with the highest accuracy reaching 98.64%, and its SESDM also performs well on Loss. In Table 1, we used the SESDM networks with different structures of U-Net network to estimate eye openness on the test set, and calculated the error between the predicted value and the true value, which proved again that when the number of floors of U-Net is 4, and channels on the 1st floor is 16, the SESDM model has the best performance for eye openness estimation.

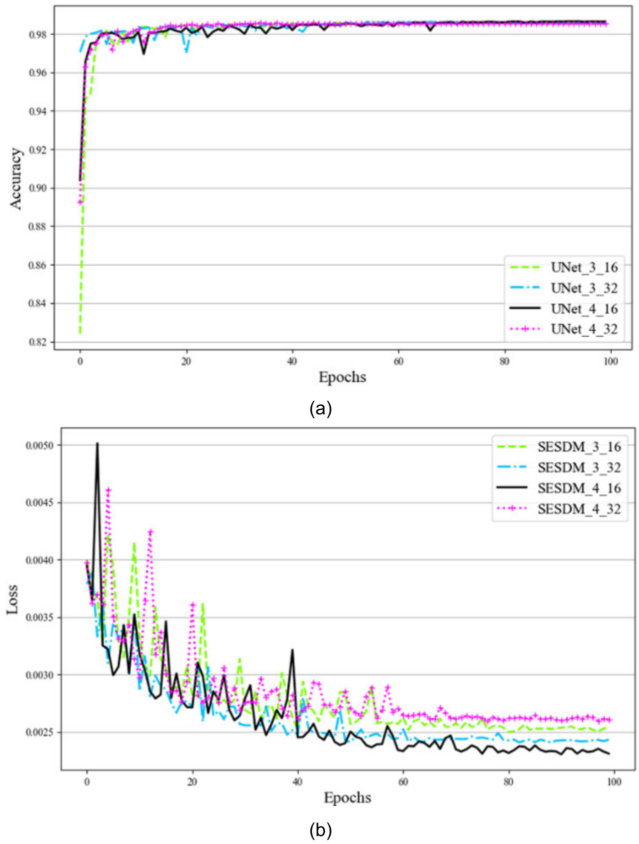
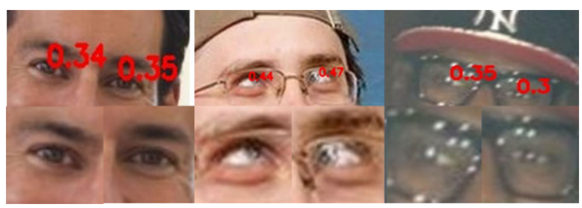


FIGURE 7. Varying curves of (a) Accuracy, and (b) Loss.



(L:0.34, R:0.36) (L:0.41, R:0.46) (L:0.33, R:0.27)

FIGURE 8. Qualitative analysis of the level of eye-openness: red numbers denote the predicted eye-openness level, the numbers under the figures denote the label of the eye-openness level.

B. COMPARING WITH OTHER METHODS FOR EYE OPENNESS ESTIMATION

When the driver is tired, blinking becomes slower and lasts longer. Therefore, by identifying the eye status of each frame, PERCLOS can effectively detect fatigue. And the pupil and iris are the most characteristic features of the eye status. Reddy *et al.* [9] proposed a multi-stream Driver Drowsiness Detection Network (DDDN), which obtained the state-of-the-art results on popular benchmarks. On the basis of Ref. [9], Wang *et al.* [11] presented a dual-stream consistency detection system based on deep learning for fatigue detection. To prove the superiority of our method in eye openness estimation, we reproduce the methods in Refs. [9] and [11] and set the same parameters on the "EYE DATA" dataset. The eye images on the test set were identified by the three models. The experimental results were shown in Table 2. The average error is calculated between the estimated values and the ground truths of eye openness levels. It can be

FIGURE 9. Sample image from the National Tsing Hua University-Driver Drowsiness Detection (NTHU-DDD) dataset.

TABLE 2. Comparison of eye openness estimation performance between different models on EYE DATA test set.

MODELS	DDDN [9]	GP-BCNN [11]	SESDM (Our)
Mean Error	0.076	0.053	0.036
Error <= 0.1	1404	1693	1863
Error >0.1	542	253	83
The number of samples	1946	1946	1946
Average time	3ms	5ms	11ms

seen from the table that the average errors of our SESDM model is less than 0.036. Although the prediction time of our model is longer, it is in a leading position in the average error and the number of images with error less than 0.1. Combined with the data in Table 1, our eye status detection method is obviously better than the methods in Refs. [9] and [11].

As shown in Fig. 8, these are the detection results of our method. This figure further shows that our method has a good detection effect on eye openness estimation of people with glasses or different skin colors.

C. COMPARING WITH OTHER FATIGUE DETECTION METHODS

The NTHU-DDD dataset was a dataset developed by National Tsing Hua University, which was used at the Asian Conference on Computer Vision Workshop on Driver Drowsiness Detection [33]. The entire dataset contains 36 subjects of different ethnicities, which were recorded with and without wearing glasses under the simulated driving conditions during a variety of daytime and nighttime. The dataset contained a lot of normal, drowsy, talking, and yawn face data in various scenarios. We select four scenarios from the NTHU-DDD dataset: BareFace, Glass, Night-BareFace and Night-Glasses. The total number of videos is 213. we randomly divide the training set, validation set and test set according to the ratio of 4:1:4. Then in order to balance the data of various scenarios, we fine-tune the data. In the end, we selected 99 videos as training set, 24 videos as the validation set and 90 videos as test set. Some screenshots from the dataset are shown in Fig. 9. Many researchers have proposed different approaches for fatigue detection and achieved good results. In order to further prove the superiority of our method, we conduct comparative experiments on the NHUT-DDD data set with the current advanced fatigue detection methods in Refs. [9], [11], [34].

The methods proposed in this article, Refs. [9] and [11] judged fatigue by detecting eye status of each frame, while the method in Ref. [34] detected fatigue via the spatial–temporal feature of driver's eyes. Therefore, we designed two schemes for the experiments. In the first scheme, we used the SESDM, DDDN [9] and GP-BCNN [11] models to identify the eye state of each frame in the NTHU-DDD test set, and calculated the PERCLOS value by formula (2). According to formula (3), when the PERCLOS value for the whole video is greater than 0.2, the video is classified as fatigue. Otherwise, it is classified as normal. In the second scheme, we reproduced FDRNet [34] on the 99 videos training set and 24 videos validation set. At test time, the predicted results are averaged across all time-steps as the final output for per clip of the test set. The results are shown in Table 3. We compare the performances of each model on the test set in terms of recall rate, precision, F1 score, wrong recognition rate, where N represents the number of wrong recognition videos. It can be seen that SESDM achieves the highest recall rate, precision, F1 score, and the lowest wrong recognition rate, which proves that the effectiveness and feasibility of SESDM. Besides, the detailed recognition results of SESDM are shown in Table 4, where the first two columns refer to data distribution of NHUT-DDD test set, and the last two columns represent the number of recognition videos. By observing the wrong recognition videos, we can find that lighting conditions and glasses have an impact on the detection of our eye

FIGURE 10. The experiment scene and PERCLOS curve, (a) while driving, (b) after parking. The blue curve represents the PERCLOS measures while driving computed using the proposed method, the red curve represents the fatigue threshold based on PERCLOS, and the green curve indicates the PERCLOS measures after parking computed by the proposed system.

TABLE 3. Comparison of fatigue detection performance between different models on NHUT-DDD test set.

Methods	Data distribution of test set	Recognition results: Fatigue	Recognition results: Normal	N	Precision, %	Recall rate, %	F1 score, %
FDRNet [34]	Fatigue (63)	54	9	13	93.10%	85.71%	89.25%
	Normal (27)	4	25				
GP-BCNN [11]	Fatigue (63)	58	5	11	90.62%	92.06%	91.34%
	Normal (27)	6	21				
DDDN [9]	Fatigue (63)	51	12	22	83.60%	80.95%	82.26%
	Normal (27)	10	17				
SESDM (Our)	Fatigue (63)	59	4	6	96.72%	93.63%	95.16%
	Normal (27)	2	25				

status. In general, our method has an excellent performance in fatigue detection.

D. FATIGUE DETECTION IN REAL DRIVING ENVIRONMENTS

In our approach, we used infrared cameras to capture facial images. Then the SESEM was used to estimate the eye openness. A shortcut between the segmentation network and the decision network was established to effectively improve the model's detection accuracy and generalization. To validate this, the experiments in real driving environments were carried out. Here, it has to be mentioned that it is not safe and not practical to detect the fatigue status of the drowsy driving on real-world roads. Fig. 10 shows the experiment scene and PERCLOS curves obtained by our SESEM model. The plots of the PERCLOS measurements against the number of frames and the fatigue threshold based on the PERCLOS were shown in the bottom of Fig. 10. In the plots of PERCLOS measurements, the blue curve represents the PERCLOS measures while driving computed using the proposed method, the red curve represents the fatigue threshold based on PERCLOS, and the green curve indicates the PERCLOS measures after parking computed by the proposed system. When the car is moving, at #1000 and #12000 frames, the strong light makes the glasses reflect strongly, which leads to the wrong result of eye status detection. At this time, the PERCLOS value has an increasing trend. During the detection, the PERCLOS values are always below 0.2, which indicates the driver is in a normal status. After parking, we increase the times and durations of eye closures to simulate the fatigue status at the #1 and #10015 frames. At the meantime, the PERCLOS

TABLE 4. Detailed recognition results of SESDM on NHUT-DDD test set.

Real data distribution of test set		Number of wrong recognition videos	
		Fatigue	Normal
Bareface	Fatigue (18)	18	0
	Normal (8)	0	8
Glasses	Fatigue (11)	10	1
	Normal (5)	0	5
Night-bareface	Fatigue (22)	21	1
	Normal (9)	1	8
Night-glasses	Fatigue (12)	10	2
	Normal (5)	1	4

value starts to increase. At the #4615 and #8889 frames, we blink at a normal frequency. The PERCLOS value became smaller, which fulfills our expectations. Although the lighting conditions have an impact on the performance of the model, the error is still acceptable. To draw a conclusion, our method performs effectively in the natural environment.

VI. CONCLUSION

In this article, a novel fatigue detection method based on the measure of eye openness has been presented, which consists of the two steps including pupil/iris segmentation and fatigue classification. This two-step model has better performances than an end-to-end fatigue detection because there are many features focusing on the pupil and iris. Specifically, the Dlib face key point detection algorithm was adopted to obtain the eye image as the experimental data. And the segmentation network performs pixel-level classification of eye images and extracts pupil and iris features in the image. Then, the shortcut of the segmentation output map is introduced to accelerate convergence and improve the detection accuracy of eye openness estimation in decision network. Finally, a PERCLOS value on the continuous detection of eye openness is computed to predict driver's state, i.e., normal or fatigue driving state. Extensive experiments demonstrated the effectiveness of our architecture, which achieves state-of-the-art performance in fatigue detection. To realize the fatigue detection on the real roads, fusing the detection of the eye openness and the head posture is also a good choice in the future work.

REFERENCES

[1] Z. Mardi, S. N. M. Ashtiani, and M. Mikaili, "EEG-based drowsiness detection for safe driving using chaotic features and statistical tests," *J. Med. Signals Sens.*, vol. 1, no. 2, pp. 130–137, Aug. 2011.

[2] M. Patel, S. K. L. Lal, D. Kavanagh, and P. Rossiter, "Applying neural network analysis on heart rate variability data to assess driver fatigue," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7235–7242, Jun. 2011.

[3] P. M. Forsman, B. J. Vila, R. A. Short, C. G. Mott, and H. P. A. Van Dongen, "Efficient driver drowsiness detection at moderate levels of drowsiness," *Accident Anal. Prevention*, vol. 50, pp. 341–350, Jan. 2013.

[4] Y. Liang, M. L. Reyes, and J. D. Lee, "Real-time detection of driver cognitive distraction using support vector machines," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 2, pp. 340–350, Jun. 2007.

[5] Y. Ji, S. Wang, Y. Zhao, J. Wei, and Y. Lu, "Fatigue state detection based on multi-index fusion and state recognition network," *IEEE Access*, vol. 7, pp. 64136–64147, 2019.

[6] A. Dasgupta, A. George, S. L. Happy, and A. Routray, "A vision-based system for monitoring the loss of attention in automotive drivers," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1825–1838, Dec. 2013.

[7] R. Knipling and P. Rau, "PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance," *Federal Highway Admin.*, vol. 31, pp. 1237–1252, Oct. 1998.

[8] M. Ramzan, H. U. Khan, S. M. Awan, A. Ismail, M. Ilyas, and A. Mahmood, "A survey on State-of-the-Art drowsiness detection techniques," *IEEE Access*, vol. 7, pp. 61904–61919, 2019.

[9] B. Reddy, Y.-H. Kim, S. Yun, C. Seo, and J. Jang, "Real-time driver drowsiness detection for embedded system using model compression of deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 438–445.

[10] 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. (CMVIT)*, Feb. 2017, pp. 105–110.

[11] Y. Wang, R. Huang, and L. Guo, "Eye gaze pattern analysis for fatigue detection based on GP-BCNN with ESM," *Pattern Recognit. Lett.*, vol. 123, pp. 61–74, May 2019.

[12] Y. Kurylyak, F. Lamonaca, and G. Mirabelli, "Detection of the eye blinks for human's fatigue monitoring," in *Proc. IEEE Int. Symp. Med. Meas. Appl.*, May 2012, pp. 91–94.

[13] M. Wang, L. Guo, and W.-Y. Chen, "Blink detection using AdaBoost and contour circle for fatigue recognition," *Comput. Electr. Eng.*, vol. 58, pp. 502–512, Feb. 2017.

[14] F. Yutian, H. Dexuan, and N. Pingqiang, "A combined eye states identification method for detection of driver fatigue," in *Proc. IET Int. Commun. Conf. Wireless Mobile Comput. (CCWMC)*, 2009, pp. 217–220.

[15] D. González-Ortega, F. J. Díaz-Pernas, M. Antón-Rodríguez, M. Martínez-Zarzuela, and J. F. Díez-Higuera, "Real-time vision-based eye state detection for driver alertness monitoring," *Pattern Anal. Appl.*, vol. 16, no. 3, pp. 285–306, Aug. 2013.

[16] B. D. Eddine, F. N. dos Santos, B. Boulebtateche, and S. Bensaoula, "EyeLSD a robust approach for eye localization and state detection," *J. Signal Process. Syst.*, vol. 90, no. 1, pp. 99–125, Jan. 2018.

[17] W. Han, Y. Yang, G.-B. Huang, O. Sourina, F. Klanner, and C. Denk, "Driver drowsiness detection based on novel eye openness recognition method and unsupervised feature learning," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2015, pp. 1470–1475.

[18] L. Wu, Y. Wang, J. Gao, and X. Li, "Deep adaptive feature embedding with local sample distributions for person re-identification," *Pattern Recognit.*, vol. 73, pp. 275–288, Jan. 2018.

[19] L. Wu, Y. Wang, and L. Shao, "Cycle-consistent deep generative hashing for cross-modal retrieval," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1602–1612, Apr. 2019.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 2, 2012, pp. 1097–1105.

[21] S. Y. Han, Y. Kim, S. H. Lee, and N. I. Cho, "Pupil center detection based on the UNet for the user interaction in VR and AR environments," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces (VR)*, Mar. 2019, pp. 958–959.

[22] J. F. May and C. L. Baldwin, "Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 12, no. 3, pp. 218–224, May 2009.

[23] A. Sahayadhas, K. Sundaraj, and M. Murugappan, "Detecting driver drowsiness based on sensors: A review," *Sensors*, vol. 12, no. 12, pp. 16937–16953, 2012.

[24] M.-H. Sigari, M.-R. Pourshahabi, M. Soryani, and M. Fathy, "A review on driver face monitoring systems for fatigue and distraction detection," *Int. J. Adv. Sci. Technol.*, vol. 64, pp. 73–100, Mar. 2014.

[25] D. Dinges, M. Mallis, G. Maislin, and J. Powell, "Evaluation of techniques for ocular measurement as an index of fatigue and as the basis for alertness management," Nat. Highway Traffic Safety Admin. (NHTSA), Washington, DC, USA, Tech. Rep. HS-808 762, 1998.

[26] B. Mandal, L. Li, G. S. Wang, and J. Lin, "Towards detection of bus driver fatigue based on robust visual analysis of eye state," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 3, pp. 545–557, Mar. 2017.

[27] P. Viola and M. Jones, "Robust real-time object detection," *Int. J. Comput. Vis.*, vol. 4, no. 4, pp. 34–47, 2001.

[28] W. Wu, C. Qian, S. Yang, Q. Wang, Y. Cai, and Q. Zhou, "Look at boundary: A boundary-aware face alignment algorithm," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2129–2138.

[29] C. Sagonas, E. Antonakos, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 faces in-the-wild challenge: Database and results," *Image Vis. Comput.*, vol. 47, pp. 3–18, Mar. 2016.

[30] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 faces in-the-Wild challenge: The first facial landmark localization challenge," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Dec. 2013, pp. 397–403.

[31] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "A semi-automatic methodology for facial landmark annotation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2013, pp. 896–903.

[32] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang, "Interactive facial feature localization," in *Proc. Eur. Conf. Comput. Vis.*, in Lecture Notes in Computer Science, vol. 7574, no. 3, 2012, pp. 679–692.

[33] C.-H. Weng, Y.-H. Lai, and S.-H. Lai, "Driver drowsiness detection via a hierarchical temporal deep belief network," in *Proc. Asian Conf. Comput. Vis.*, in Lecture Notes in Computer Science, 2016, pp. 117–133.

[34] 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.

QIANYANG ZHUANG is currently pursuing the master's degree with the School of Engineering, Zhejiang Normal University, China. His research interests include deep learning and image processing.

ZHANG KEHUA received the Ph.D. degree in mechatronic engineering from the Zhejiang University of Technology (ZJUT). He is currently an Associate Professor with ZJNU, where he serves the Key Laboratory of Urban Rail Transit Intelligent Operation and Maintenance Technology and Equipment of Zhejiang Province. His current research interests include computer vision, deep learning, and image processing.

JIAYI WANG is currently pursuing the bachelor's degree with the College of Engineering, Zhejiang Normal University, China. Her research interests include deep learning and image processing.

QIANQIAN CHEN is currently pursuing the bachelor's degree with the College of Engineering, Zhejiang Normal University, China. Her research interests include deep learning and image processing.