A Fatigue Driving Detection Algorithm Based on Facial Multi-Feature Fusion

KENING LI[1,3], YUNBO GONG[2], AND ZILIANG REN[4,5], (Member, IEEE)
[1]School of Traffic and Environment, Shenzhen Institute of Information Technology, Shenzhen 518172, China
[2]School of Civil Engineering and Transportation, South China University of Technology, Guangzhou 510640, China
[3]Guangdong Key Laboratory of Intelligent Transportation System, School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou 510275, China
[4]CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[5]Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong

Corresponding author: Ziliang Ren (zl.ren@siat.ac.cn)

ABSTRACT Researches on machine vision-based driver fatigue detection algorithm have improved traffic safety significantly. Generally, many algorithms do not analyze driving state from driver characteristics. It results in some inaccuracy. The paper proposes a fatigue driving detection algorithm based on facial multi-feature fusion combining driver characteristics. First, we introduce an improved YOLOv3-tiny convolutional neural network to capture the facial regions under complex driving conditions, eliminating the inaccuracy and affections caused by artificial feature extraction. Second, on the basis of the Dlib toolkit, we introduce the Eye Feature Vector(EFV) and Mouth Feature Vector(MFV), which are the evaluation parameters of the driver's eye state and mouth state, respectively. Then, the driver identity information library is constructed by offline training, including driver eye state classifier library, driver mouth state classifier library, and driver biometric library. Finally, we construct the driver identity verification model and the driver fatigue assessment model by online assessment. After passing the identity verification, calculate the driver's closed eyes time, blink frequency and yawn frequency to evaluate the driver's fatigue state. In simulated driving applications, our algorithm detects the fatigue state at a speed of over 20fps with an accuracy of 95.10%.

INDEX TERMS Traffic safety and environment, fatigue driving detection, machine vision, convolutional neural network.

I. INTRODUCTION

With the rapid growth of the number of cars, there are more and more traffic accidents, which brings huge potential safety hazards to travel. In order to minimize the occurrence of traffic accidents, recently, the government has introduced multiple related policies, and achieved significant results. However, at this stage, traffic accidents are still one of the main threats to life safety. For example, lack of road safety driving awareness, drunk driving, and fatigue driving are the main factors that cause traffic accidents. Among them, fatigue driving accounts for 14% -20% of the causes of traffic accidents, about 43% in heavy traffic accidents, and about 37% in traffic accidents on large trucks and highways [1].

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

According to the NHTSA (National Highway Traffic Safety Administration) survey of driving vehicles, more than 70% of the surveyed drivers experienced fatigue driving [2], [3]. The NTSB (The National Transportation Safety Board) of the United States found that after investigating 120 accidents related to drivers, nearly 60% of them were related to driver fatigue. In France, the number of casualties caused by driver fatigue is 23% of the total. Australia NRSA (National Road Safety Administration) released a report on the cause of major traffic accidents. It shows that fatigue driving accounted for more than 15%. After investigating the causes of traffic accidents, Flatley [4] found that more than 21.6% of traffic accidents are related to fatigue driving. In addition, the investigation also found that other traffic accidents caused by improper operation and carelessness are also related to fatigue driving to some extent.

The Road Traffic Safety Law of China clearly states that if drivers drive for more than 4 hours without rest, he will be considered to be fatigued driving [5], [6]. If the driver has excessive fatigue driving behavior, the traffic control department can conduct it penalties and deduction of driving license. Although the regulation can reduce the driver's excessive fatigue driving behavior to a certain extent, giving fatigue warnings at critical times can greatly reduce the occurrence of traffic accidents caused by fatigue driving. Especially for drivers engaged in long-distance passenger transportation and freight transportation, they need to drive motor vehicles continuously for a long time due to work requirements. However, it is difficult to maintain a high alert state all the time while driving a vehicle. Therefore, real-time detection and alarm of fatigue status is even more important.

At present, the detection of fatigue driving is mainly divided into subjective method and objective method [7], [8]. The subjective method is based on a questionnaire survey. Well-known questionnaires include the Stanford Sleep Scale, Pearson Fatigue Scale, Driver Record Form, and Cooper-Harper Evaluation Questionnaire, which include subjective load assessment,sleep habits table and so on. The questionnaire survey is based on the driver's subjective thinking to answer the questions in the questionnaire. It has a strong subjectivity, so it cannot be used as a standard method for detecting fatigued driving.

The objective method is to use the auxiliary tools to detect the driver's physiological characteristics or monitor the vehicle information, etc., and to judge fatigue driving [9], [10]. Mainly divided into three categories:

◇ Fatigue detection based on physiological characteristics

Studies have shown that as the degree of fatigue increases, the physiological indicators of the human body will gradually deviate from the normal value [11].Therefore, fatigue can be judged according to the change of the driver's physiological characteristics. Common features include EEG, ECG, and EMG. Among them, EEG is regarded as the "gold standard" for detecting fatigue [12].

◇ Fatigue detection based on vehicle behavior characteristics

This type of detection method mainly collects and analyzes the relevant information of the vehicle itself during the driving process of the vehicle, to determine whether the driver of the vehicle is fatigued [13], [14].

◇ Fatigue detection based on facial features

When the human body is in a fatigue state and a non-fatigue state, part of its body parts will show very different performances [15], [16]. For example, in the fatigue state, the human body may experience head droop, body tilt, increased blink frequency, and yawning.

1. Use the position of the head to determine fatigue: when the driver is tired, his head may appear tilted, swaying, etc., so the driver's head movement can be detected to determine whether he is fatigued.

2. Use the state of the eyes to determine fatigue: detect fatigue through eye characteristics such as blinking frequency and eye closing time. After a large number of experiments and verifications, the Carnegie Mellon Institute in the United States proposed the PERCLOS method for measuring fatigue. The method is considered to be the most reliable and effective fatigue determination method at present. The driver 's eye closure can be obtained based on this parameter to judge its fatigue state [17].

3. Fatigue is judged by the degree of opening and closing of the mouth: the fatigue is judged according to the different performance of the driver's mouth state when speaking normally and yawning. First obtain the image of the mouth through the video capture device, and then import the opening and closing features of the mouth Neural network system to judge fatigue based on the duration of mouth opening and closing [18].

It is a non-contact method to determine fatigue based on the driver's facial features [19], [20]. It does not cause interference and impact on the driver while driving the vehicle, and has the advantages of fast speed and strong operability. Therefore, compared with the other two types of fatigue driving detection methods, this method is currently the most concerned and widely used.

Universities, research institutes and enterprises have conducted long-term and in-depth research on fatigue driving testing. Abe *et al.* [21] obtained the best results of fatigue detection by studying the relationship between eye state and fatigue. The eye characteristics studied by them mainly include eye opening and closing, eye movement, pupil, etc. Devi and Bajaj [22] proposed a fatigue detection model, in which the system first locates the face of the driver image, and then extracts the mouth and eye features in the localized face area. Finally, through the fatigue detection system, the features are comprehensively processed to determine whether the driver is fatigued. By studying the positioning and tracking of the eyes under infrared light at night, Singh *et al.* [23] analyzed the relationship between the eye's state and fatigue. Haro *et al.* [24] proposed the use of infrared light for eye positioning and Kalman filtering for eye tracking, which is highly robust. Coetzer and Hancke [25] found that the Adaboost algorithm has certain advantages in some aspects of face detection, after comparing the Adaboost algorithm with ANN and SVM.

Although the technology of fatigue detection has made better progress and results, it still need to be improved.

◇Detection methods based on physiology and behavior usually require the driver to wear or additionally install more physiological information monitoring devices, which affects the comfort of the driver's normal driving. Moreover, equipment that collects physiological information is often expensive and vulnerable, which is not conducive to the popularization of fatigue driving detection systems.

◇The detection method based on vision usually uses Adaboost Classifier Algorithm for face localization [26], [27]. However, when the driver wears glasses or sunglasses, light changes, and the face is partially occluded, Adaboost cannot accurately locate the face position and promptly warn of fatigue driving.

◇At present, the common algorithms judge fatigue by state of the driver's eyes and mouth. However, these algorithms do not take driver's individual characteristics into account. In fact, the algorithm has high misjudgment, using a fixed threshold to determine the state of the eyes and mouth.

As discussed above literatures, results of the driving fatigue detection have defect of high intrusion, low robustness, and low reliability. Therefore, we propose a new algorithm. The innovations are as follows:

◇We design a driver's face detection architecture based on the improved YOLOv3-tiny convolutional neural network, and trains the network with the open-source dataset WIDER FACE [28]. Compared with other deep learning algorithms, such as YOLOv3 [29], MTCNN [30], the algorithm based on the improved YOLOv3-tiny network improve the face recognition accuracy, simplify the network structure, and reduce the amount of calculation. And then, it is more convenient to transplant to the mobile.

◇Most of the existing algorithms are based on the PER-CLOS, which uses the driver's eyes state as a feature to judge fatigue. In fact, when the driver's eyes are too small, the algorithm has high misjudgment. Similarly, the algorithm based on yawn frequency is also related to the size of the driver's mouth [31], [32]. Therefore, we design the eye and mouth SVM classifier, which takes driver characteristics into account driver characteristics. It judges fatigue based on the actual driver's eyes and mouth size. It has high accuracy.

◇The existing machine learning algorithms that consider individual characteristics often train classifiers by initialization before the system starts, which requires re-initialization every time the driver is changed. Not only is it a waste of time, it also does not ensure that every initialization works well. Therefore, we constructed the driver identity information library. There are three types of driver identity information in the system: driver biometrics, driver eye state classifier, and driver mouth state classifier. We train classifiers in advance and store them into the driver identity information library. Then, through identity verification, the driver's classifiers are called before system startup. It not only simplifies the initialization, but also avoids inaccuracies due to entering the identity manually.

This paper is divided into the following 4 parts.

The first chapter is the introduction, which mainly introduces the background and research significance of our fatigue driving detection system, and briefly expounds the research status of domestic and foreign fatigue driving detection. Moreover, according to the shortcomings of current research, we propose a new algorithm. Finally, we introduce the innovation of our algorithm.

The second chapter is the introduction of the algorithm. Firstly, we use the improved YOLOv3-tiny network for face detection. Secondly, we introduce how to combine the Dlib toolkit to extract facial feature parameters. Then, it is introduced how to establish the driver identity information library. Finally, we introduce how to construct the driver's identity

verification model and the fatigue assessment model, and how to use the model to judge the fatigue state.

The third chapter is the experimental analysis. Firstly, the experimental environment and data set are introduced. Then we use qualitative description and quantitative evaluation to measure face detection and feature point location. Finally, we evaluate our fatigue driving detection algorithm in two directions: accuracy and real-time.

The fourth chapter is the conclusion, which mainly summarizes the main work content of this paper, analyzes the shortcomings of the system and the aspects that need to be improved. And then, we propose the future optimization direction and prospect of the algorithm.

II. METHODOLOGY

As shown in Figure 1, our algorithm includes the following 3 modules.

Identity Entry: Firstly, we use the camera to collect driver biometric images, eye classification images and mouth classification images. Based on deep learning theory, we apply the improved YOLOv3-tiny network to locate suspected face regions from complex backgrounds. Secondly, according to the driver's face regions coordinates, the Dlib toolkit is used to extract facial feature points coordinates, by which we calculate 128-dimensional Feature Vector, Eye Feature Vector (EFV), and Mouth Feature Vector (MFV) of the driver's face in the image. Then, to get the eye state classifier that takes driver characteristics into account, support vector machines (SVM) are trained with the eye feature vector in the open-eye image and the closed-eye image. The same goes for the mouth. Finally, the driver's biometric, eye state classifier, and mouth state classifier are stored in the driver's identity information library.

Identity Verification: Firstly, we use the camera to collect images including driver biometric, and based on deep learning theory, we apply the improved YOLOv3-tiny network to locate suspected face regions from complex backgrounds. Secondly, according to the driver's face regions coordinates, the Dlib toolkit is used to extract facial feature points coordinates, by which we calculate 128-dimensional Feature Vector of the driver's face in the image. Then, it is compared with all the stored driver biometric in the driver identity information library. Finally, according to the comparison result, the eye classifier and mouth classifier of the corresponding driver are called for online recognition.

Online Recognition: The original data source is the real-time camera video. Firstly, based on deep learning theory, we apply the improved YOLOv3-tiny network to extract suspected face regions from complex backgrounds. Secondly, according to the driver's face regions coordinates, we use the Dlib toolkit extract the driver's eye and mouth coordinates, by which we calculate the driver's EFV and MFV in real time during driving. Then, according to the EFV, we use the eye state classifier obtained by identity verification to judge the driver's eye state. The same goes for the mouth. Finally, based on the eye state and mouth state of each picture detected
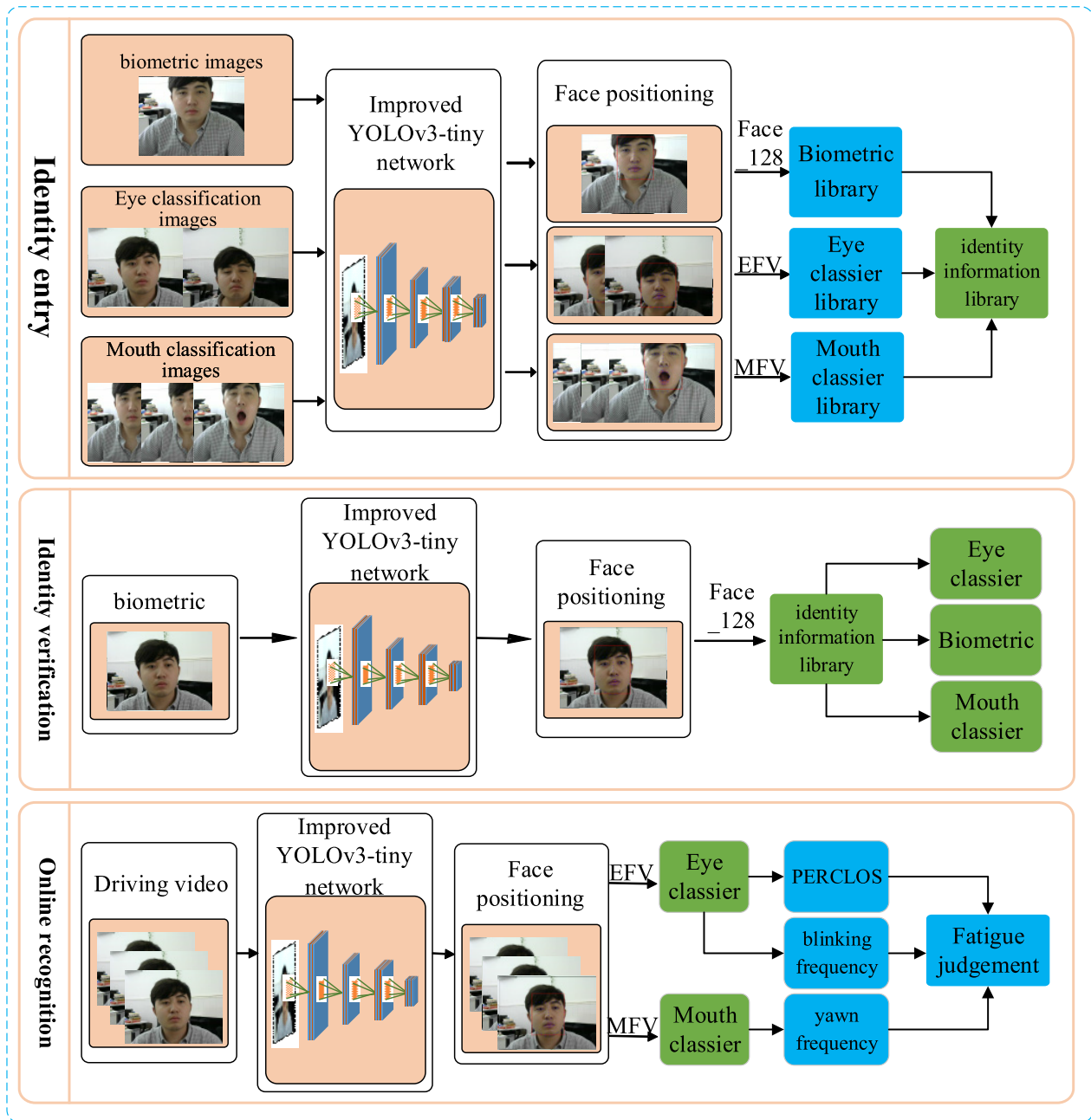
FIGURE 1. Algorithm structure diagram.

over a period of time, we calculate the driver's PERCLOS, blinking frequency, and yawn frequency to judge the driver's fatigue state.

A. FACE DETECTION BASED ON THE IMPROVED YOLOv3-TINY NETWORK

The correctness of the face detection directly affects the performance of the driving fatigue detection algorithm. So, accurate and rapid face detection is the fundamental task of the driving fatigue detection algorithm. In the traditional algorithm, Viola and Jones [33] proposed that the Haar-Like [34]

features of the input image can be extracted. Based on these extracted features, AdaBoost algorithm is used to train multiple weak classifiers. Finally, multiple weak classifiers are cascaded to obtain a strong classifier, which is the final face detector. This method effectively improves the performance of face detection, and it is still used and improved to this day. Recently, with the continuous development and application of deep learning, it provides a new method for face detection and segmentation [35]. It can be divided into two categories: One is a multi-level detection algorithm based on proposal region. The second is the target detection algorithm based on

the anchor box. The former's representative algorithms are Faster-Rcnn [36] and MTCNN [30]. The latter's representative algorithms are S3FD [37], SSH [38]. Compared with traditional methods [39], face detection based on convolutional neural network(CNN) avoids the artificial extraction of features. With the support of data sets, face detection performance has been greatly improved.

YOLO [40] is You Only Look Once, which means you only need to look at the picture once to get the target information. YOLO treats target detection as a regression problem to solve, and uses an end-to-end convolutional neural network to extract the characteristics of the input image. It can obtain the position, size and category information of the target in the image.

YOLOv3 is an improved version of the YOLO algorithm. And it is one of the best algorithms in the field of target detection. Based on YOLO, YOLOv3 refers to many excellent research results. When using YOLOv3 to detect 320×320 images, the detection accuracy is consistent with the SSD algorithm, but it is three times faster than the SSD algorithm.

YOLOv3-tiny is a lightweight target detection model based on YOLOv3. When detecting images on Pascal Titan X, the detection speed can reach 220FPS, and is far higher than the general network. The YOLOv3-tiny algorithm has the following advantages: 1) Fast detection speed. The detection result can be obtained by running a neural network once for each test image, and it can be used for real-time detection; 2) Global understanding of the image. The information around the target can be learned during training, and the background error rate is less than half of the Fast R-CNN algorithm; YOLOv3-tiny can be used on devices with low computing power such as embedded devices, but the detection accuracy of YOLOv3-tiny is greatly reduced compared to YOLOv3.

The network structure of YOLOv3-tiny is obtained by simplifying the network structure of YOLOv3.

The YOLO [40] (You Only Look Once) model is a fast target detection model based on deep learning [41], [42]. It is a separate end-to-end network that turns target detection into a regression problem. To be more specific, the method of regression and the CNN [43], [44] are used to replace the sliding window of the traditional target detection to realize the feature extraction of the driver's face. This method of feature extraction is less affected by the external environment and has the advantage of extracting target features quickly.

YOLOv3-tiny has a 23-layer network, including 13 Convolution layers, 6 Max Pooling layers, 1 Up Sampling layer, 1 Fully connected dence layer, and 2 Output layers.To simplify the network and reduce the computation, we transform the regression of multiple targets into a single target according to the regression idea of YOLO model. And then, we improve YOLOv3-tiny network to locate suspected face regions. The improved network structure is shown in Figure 2.

In the YOLOv3-tiny network training phase, we use the WIDER FACE (Face Detection Data Set and Benchmark)

(http://wider-challenge.org/2019.html) [28] data set as the driving data. The WIDER FACE dataset includes 32,203 images and 393,703 marked faces, which is one of the most common face databases. The data set includes different scales, poses, occlusions, expressions, makeup, lighting, as shown in Figure 3.

The WIDER FACE data set has the following features:
• The image resolution is generally high, and all image are color images.
• Each image has a large number of faces, and each image contains an average of 12.2 faces, with more dense small faces.
• The data set is divided into three types: training set, test set, and verification set, which respectively account for 40%, 50%, and 10% of the data set.

Firstly, based on the YOLOv3-tiny network, the picture of the WIDER FACE data set is adjusted to 10 different sizes, and the grid cells are arranged on the adjusted pictures by 13×13 and 26×26. Then, we find the location of the driver's face on the non-overlapping grid cell and classify it. For each grid cell, the network outputs B bounding boxes, corresponding confidence, and the conditional probability of the driver's face. Finally, non-maximal values are used to suppress redundant bounding boxes. The confidence formula is given as Equation (1).

$$score = P_r(Object) * IOU_{pred}^{truth} \qquad (1)$$

where $P_r(Object)$ is the probability of the driver's face. If the face is included, $P_r(Object) = 1$; otherwise $P_r(Object) = 0$. IOU_{pred}^{truth} is the intersection over union(IOU) of the bounding box to the real box.

The YOLOv3-tiny network loss function consists of the central error term of the bounding box, the width and high error term of the bounding box, the error term of the prediction confidence, and the error term of the prediction category. Based on the YOLOv3-tiny network completed by offline training, we realize the location of the driver's suspected face area and provides an accurate driver's face image for the following algorithm.

B. DRIVER'S FACIAL MOTION FEATURE EXTRACTION
1) FACE FEATURE LOCATION AND 128-DIMENSIONAL FEATURE VECTOR EXTRACTION BASED ON THE DLIB TOOLKIT

On the driver's face area located by the improved YOLOv3-tiny network, the Face keypoint detection model based on the Dlib [45] library(As is shown in Figure 4(a)) is used to extract the fine-grained features of the driver's face. The Dlib library contains 68 face key points, which uses the method of cascading shape regression to query the key points of the face component.

Dlib is a modern C ++ toolbox, which contains machine learning algorithms and tools designed in C ++, and used to solve practical problems.
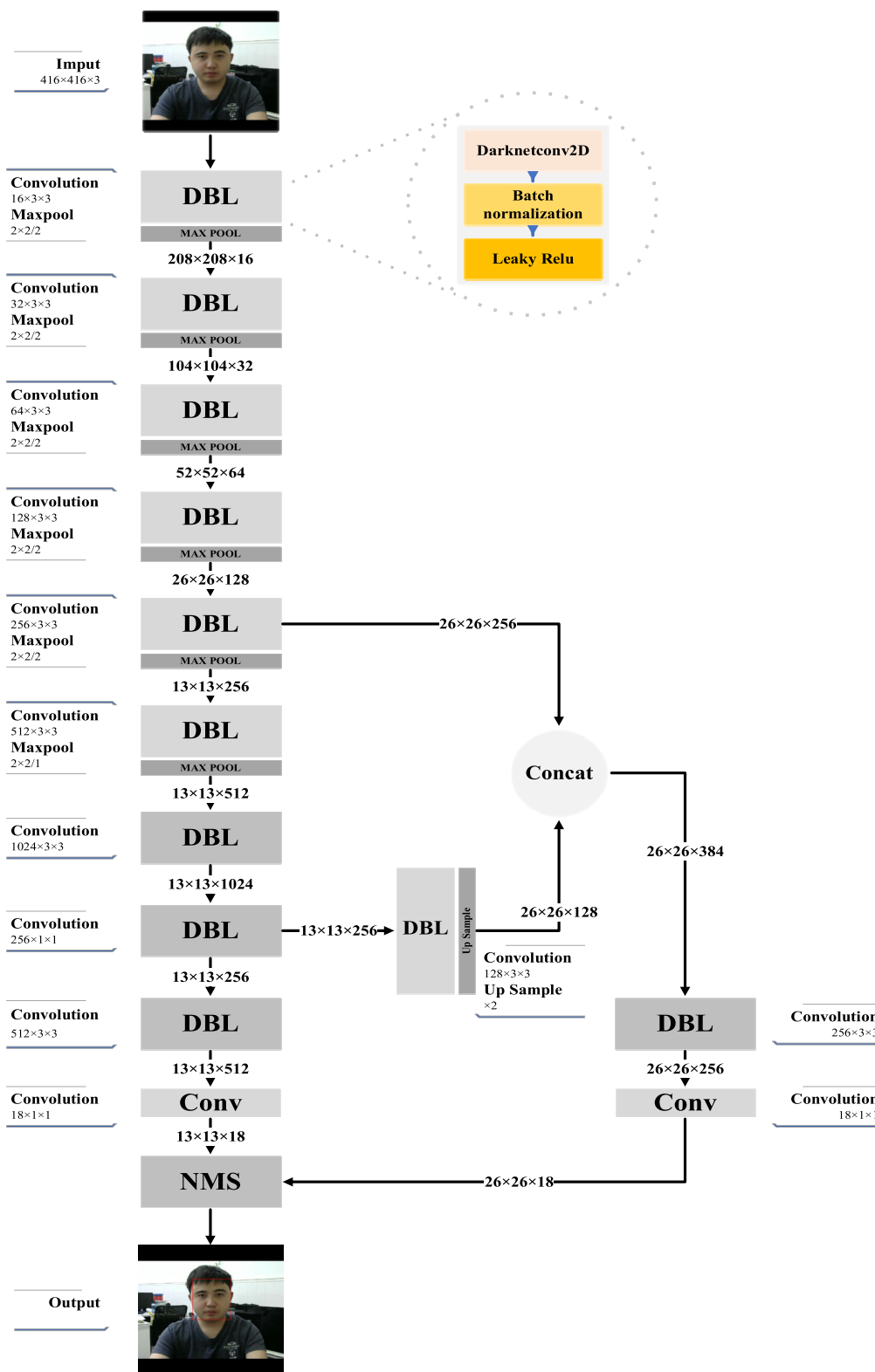
FIGURE 2. Improved YOLOv3-tiny network structure diagram.

In the face key point detection, Dlib adopts the method in [46]–[48] and provides a model trained based on millions of faces. This method uses the integration of regression trees [49]–[51] to estimate the position of facial key points directly from the sparse subset of pixel intensity, with high detection accuracy and very little time-consuming. This method will be used in the face feature extraction proposed in this paper.
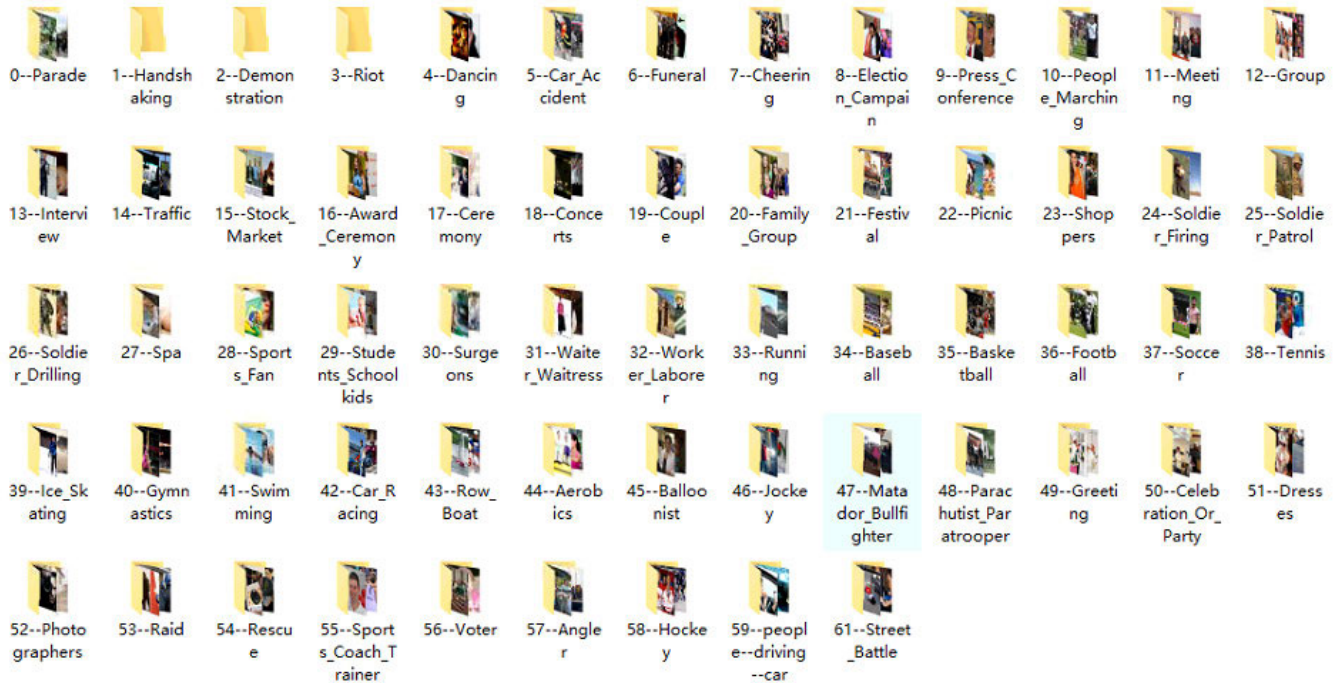
| 0--Parade | 1--Handsh aking | 2--Demon stration | 3--Riot | 4--Dancin g | 5--Car_Ac cident | 6--Funeral | 7--Cheerin g | 8--Electio n_Campai n | 9--Press_C onference | 10--Peopl e_Marchin g | 11--Meeti ng | 12--Group |

| 13--Intervi ew | 14--Traffic | 15--Stock_ Market | 16--Award _Ceremon y | 17--Cere mony | 18--Conce rts | 19--Coupl e | 20--Family _Group | 21--Festiv al | 22--Picnic | 23--Shop pers | 24--Soldie r_Firing | 25--Soldie r_Patrol |

| 26--Soldie r_Drilling | 27--Spa | 28--Sport s_Fan | 29--Stude nts_School kids | 30--Surge ons | 31--Waite r_Waitress | 32--Work er_Labore r | 33--Runni ng | 34--Baseb all | 35--Baske tball | 36--Footb all | 37--Socce r | 38--Tennis |

| 39--Ice_Sk ating | 40--Gymn astics | 41--Swim ming | 42--Car_R acing | 43--Row_ Boat | 44--Aerob ics | 45--Balloo nist | 46--Jocke y | 47--Mata dor_Bullfi ghter | 48--Parac hutist_Par atrooper | 49--Greeti ng | 50--Celeb ration_Or_ Party | 51--Dress es |

| 52--Photo graphers | 53--Raid | 54--Rescu e | 55--Sport s_Coach_T rainer | 56--Voter | 57--Angle r | 58--Hocke y | 59--peopl e--driving --car | 61--Street _Battle | | | | |

FIGURE 3. Improved YOLOv3-tiny network structure diagram.
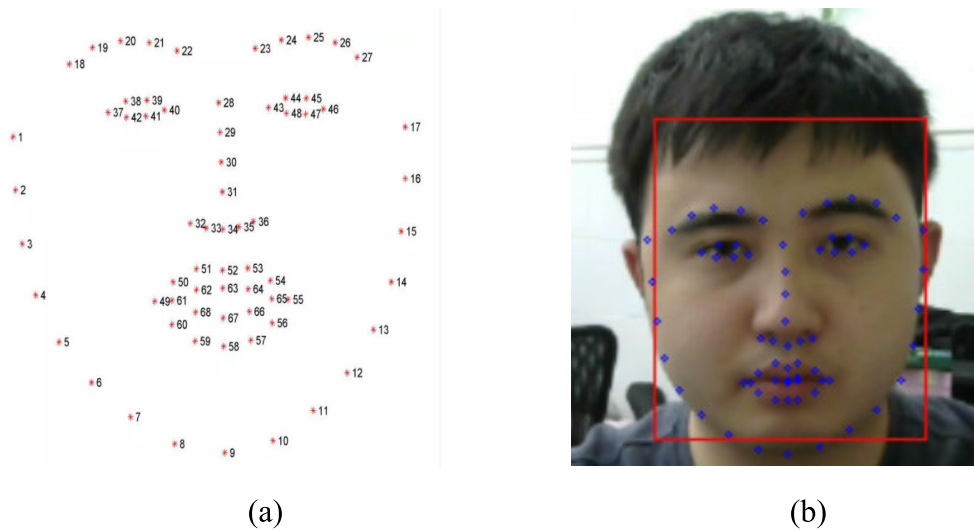

(a) (b)

FIGURE 4. Driver's face feature point acquisition based on Dlib. (a) Dlib face feature; (b) Face feature point positioning effect.

When the driver's face is detected, the feature points of the face are obtained in real time by the above algorithm, as shown in the Figure 4(b).

After extracting 68 feature points with the Dlib toolkit, they can be used to form the face information into a 128-dimensional Feature Vector [52]–[54]. In this vector space, the Euclidean distance of the same face is closer than that of different faces. Therefore, 128-dimensional Feature Vectors extracted based on the Dlib toolkit can be used as driver biometrics for identity verification.

2) EYE STATE PARAMETERS EXTRACTION BASED ON EFV

As discussed above, whether the fatigue detection algorithm based on the traditional PERCLOS or the blink frequency is dependent on the judgment of eye state. The methods mostly use the P80 standard to extract the parameters of the eye state. Firstly, the image of the eye is pre-processed through image processing. Secondly, the contour of the driver's eyes is fitted by ellipse fitting. Finally, the ratio of the major axis to the minor axis of the ellipse is used as a parameter to characterize the state of the eye. This method relies on the effect of eye
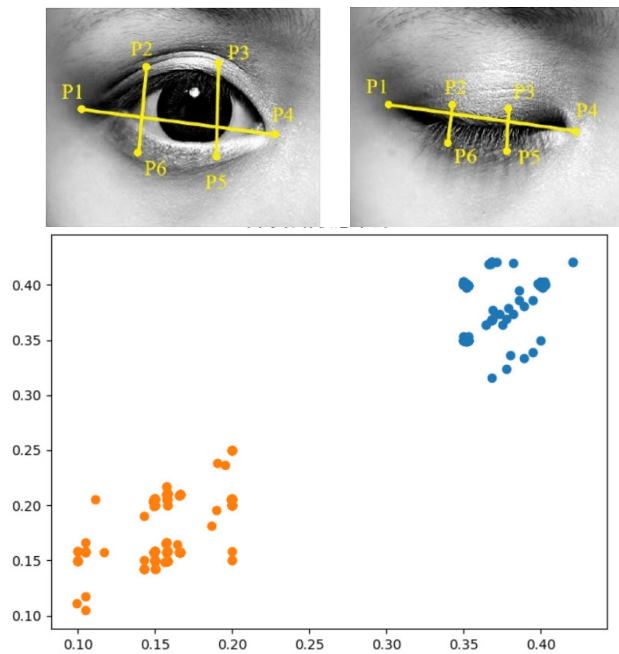
FIGURE 5. Eye state and EFV difference (Orange means closed, and blue means open).

FIGURE 6. Mouth state and MFV difference.

image preprocessing. In a real scenario, this method may have low accuracy due to constant changes in lighting conditions and the driver's head posture during driving.

To this end, based on Dlib facial feature point localization, the paper proposes a new parameter, Eye Feature Vector (EFV), which can be used to evaluate the driver's eye state. According to the Dlib eye feature points, EFV can be defined: the output of the driver eyes extraction module, to gain the parameter which can indicate the fatigue status of driver. In this module, the ellipse fitting method is applied to obtain the shape of pupils of driver. Eyes state (opening or closed) can be decided according to the relationship between the long and short axes of the ellipse. Furthermore, the fatigue status of driver is evaluated by PERCLOS.

$$EFV = (\frac{||p_2 - p_6||}{||p_1 - p_4||}, \frac{||p_3 - p_5||}{||p_1 - p_4||}) \quad (2)$$

where P_i, $i = 1, 2, \ldots, 6$ is the eye feature point coordinate.

As shown in Figure 5, when the driver's eyes are in different states, the eye feature points have significant differences. As seen in the plane scatter plot (where blue is the EFV of the open-eye picture and orange is the EFV of the closed-eye picture), when the driver's eyes are in different states, there are also significant differences in EFV, which are in line with eye feature points. Therefore, EFV can be used as a parameter to characterize the state of eyes for driver fatigue detection algorithms.

3) MOUTH STATE PARAMETERS EXTRACTION BASED ON MFV

Similar to the PERCLOS and blink frequency, the yawn frequency is also an important index for evaluating fatigue. It is inaccurate to judge fatigue based on the degree of mouth
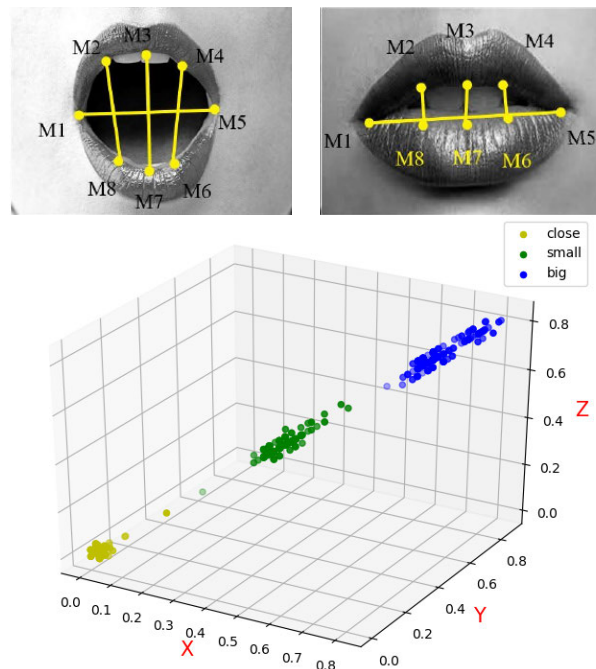
opening. In the normal driving process of the driver, speaking also appears as a change in the degree of mouth opening, which greatly interferes with fatigue judgment based on the degree of mouth opening. After analyzing the yawning process, we find that: when speaking, the mouth is opened to a small extent and its opening duration is short; In the yawn state, the mouth is opened to a greater extent and its opening duration is longer.

In order to distinguish between the yawn and speaking, the paper divides the mouth state into three types: closed mouth, small mouth, and big mouth. Similar to the eye state parameters, based on Dlib facial feature point localization, the paper proposes a new parameter, Mouth Feature Vector (MFV), which can be used to evaluate the driver's mouth state. According to the Dlib mouth feature points, MFV can be defined:

$$MFV = (\frac{||M_2 - M_8||}{||M_1 - M_5||}, \frac{||M_3 - M_7||}{||M_1 - M_5||}, \frac{||M_4 - M_6||}{||M_1 - M_5||}) \quad (3)$$

where M_i, $i = 1, 2, \ldots, 8$ is the mouth feature point coordinate.

As shown in Figure 6, when the driver's mouth is in different states, the mouth feature points have significant differences. As seen in the plane scatter plot, when the driver's mouth is in different states, there are also significant differences in MFV, which are in line with mouth feature points. Therefore, MFV can be used as a parameter to characterize the state of mouth for driver fatigue detection algorithms.

C. DRIVER IDENTITY INFORMATION LIBRARY
1) DRIVER EYE STATE CLASSIFIER LIBRARY

As mentioned above, traditional driver fatigue detection algorithms are mostly based on the P80 criterion, which uses a
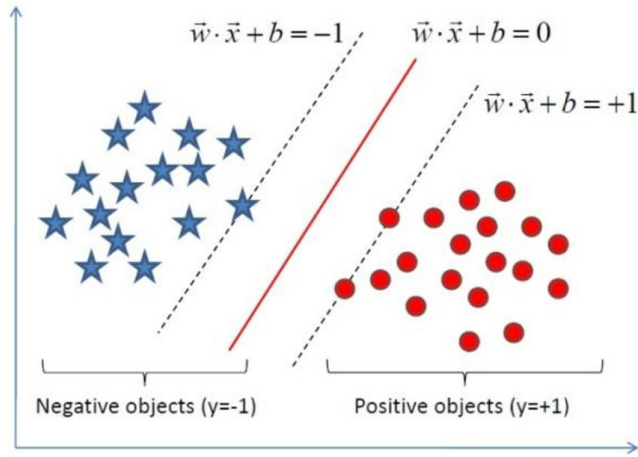
FIGURE 7. The optimal hyperplane of SVM.

fixed threshold to judge the driver's eye state without considering driver characteristics. To this end, we establish the driver eye state classifier library. According to different driver characteristics, we collect images of the driver's different eye states within a specific time, calculate EFV in different states, and train the SVM [55] classifier to judge the driver's eye state.

For a binary classification problem, if there are classified data samples $\{(x_1, y_1), \ldots, (x_i, y_i), \ldots, (x_m, y_m)\}, i = 1 \ldots m$, where m represents sample data and $x_i \in R^n$ represents for m-dimensional data, the corresponding classification label $y_i \in (-1, 1)$. In the case of linear separability, based on the soft-separation maximization criterion, the SVM algorithm seeks an optimal hyperplane to separate the two types of data samples, where the distance between the sample point closest to the hyperplane and the hyperplane is the largest.

As shown in Figure 7, a schematic diagram of finding the optimal hyperplane for a two-dimensional space. The points of triangles and circles represent two types of data, H1 and H2 are the boundaries of the two classes parallel to the optimal hyperplane. They are determined by the sample points of the closest points to the optimal hyperplane in each category, and the distance from the boundary to the optimal hyperplane is called the classification interval $M \arg in = 1/||w||$. As can be seen from the figure, the classification interval between the two types of samples is $2*M \arg in = 2/||w||$, where the optimal classification hyperplane can be expressed as Equation (4):

$$w^{*T} \cdot x + b^* = 0 \tag{4}$$

The normal vector wT and the intercept b determine the superclass surface function. The constrained optimization problem can be defined:

$$\begin{cases} \min_{w,b} J(w) = \dfrac{1}{2} \|w\|_2^2 \\ s.t.\ y_i(w^T \cdot x_i + b) \geq \quad 1,\ i = 1, 2, \ldots N \end{cases} \tag{5}$$

In the offline training phase of the driver's eye, the improved YOLOv3-tiny is used to detect the face. Based on face feature points, the driver's EFV is calculated to form the training set. Among them, When $y_i = +1$, x_i is a positive sample, indicating that the eye of driver is open, When $y_i = -1$, x_i is a negative sample, indicating that he eye of driver is close. Combined with the constraints of Equation(5), the hyperplane parameters w^T and b can be solved to construct the driver eye state classifier library.

2) DRIVER MOUTH STATE CLASSIFIER LIBRARY
Similar to the eye, traditional algorithms use a fixed threshold to judge the driver's mouth state without considering driver characteristics. To this end, we establish the driver mouth state classifier library. According to different driver characteristics, we collect images of the driver's different mouth states within a specific time, calculate MFV in different states, and train the SVM classifier to judge the driver's mouth state.

The mouth state classification is not a binary classification problem. However, SVM is a binary classifier. The construction of SVM multi-class classifiers is mainly through the indirect method, that is, combine multiple two classifiers to construct multi-classifiers.

So, in the offline training phase of the driver's mouth, the improved YOLOv3-tiny is used to detect the face. Based on face feature points, the driver's MFV is calculated to form the training set. Then, two SVM classifiers are constructed. The first classifier is trained to judge the open-closed mouth state, and the second classifier is trained to judge the small-big mouth state, as shown in Figure 8, to build the driver mouth state classifier library.
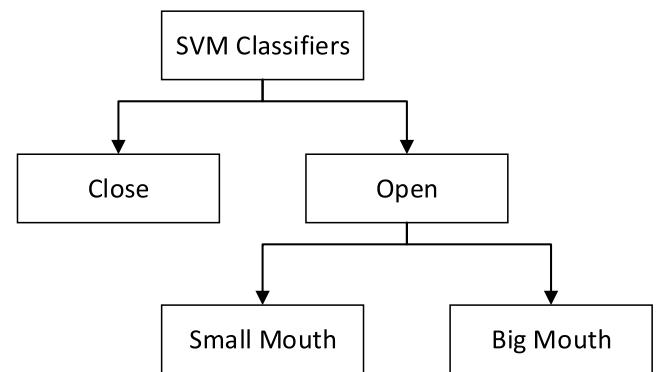

FIGURE 8. Mouth state classifier model.

3) DRIVER BIOMETRIC LIBRARY
As mentioned above, considering driver's characteristics can improve detection accuracy. However, considering that offline training needs to be performed again, every time the driver is changed, we have established the driver biometric library based on section 2.2.1. Every driver only needs to input his eye state classifier, mouth state classifier and biometric to the driver identity information library once. Before driving, the system will extract the biometric of the current

driver and compare it with the driver biometric information library. If successful, the driver's eye state classifier and mouth state classifier are called for online recognition; if unsuccessful, the driver will be reminded to input his own identity information into the driver identity information library.

D. DRIVER'S IDENTITY VERIFICATION MODEL

According to the driver's identity information library, before assessing the fatigue, the driver needs to complete identity information verification. First we use the camera to collect biometric information of the current driver, that is, the 128-dimensional feature vector of the driver's face. Then, calculate the Euclidean distance between it and all 128-dimensional feature vectors in the driver's biometric library. The calculation formula is:

$$d = \sqrt{\sum_{i=1}^{128}(x_i - y_i)^2} \tag{6}$$

where x_i is the i th dimension of the 128-dimensional feature vector currently collected, and y_i is the i th dimension of the 128-dimensional feature vector in the driver's biometric library.

We use 0.6 as the system's decision threshold. When all d values are greater than 0.6, it is determined that the current driver is not in the driver biometric library, that is, the verification fails; When there is a d value less than 0.6, it is determined that the current driver is in the driver's biometric library, that is, the verification is passed. Then find the identity information corresponding to the minimum d value as the result of the identity verification, that is, the driver's eye and mouth state classifier is called for detection before the system starts.

E. DRIVER FATIGUE ASSESSMENT MODEL
1) FATIGUE JUDGMENT BASED ON PERCLOS

Driver fatigue is a description of the state, and its corresponding fatigue level is a dynamically changing process. Carnegie Mellon Research Center Wierwille proposed Percentage of Eye Closure (PERCLOS). It has been widely accepted and adopted by many researchers as an effective indicator of fatigue driving. PERCLOS [56] is a physical quantity that measures the state of human fatigue (drowsiness), which is defined as the time taken by the eyes to be closed per unit time. The U.S. Federal Highway Administration and the National Highway Traffic Safety Administration simulated driving in a laboratory, which has verified the effectiveness of PERCLOS in characterizing driver fatigue. PERCLOS is defined as:

$$PERCLOS = \frac{N_{close}}{N_{total}} \times 100\% \tag{7}$$

where N_{close} is the number of closed eyes images in a specific time, and N_{total} is the total number of images in a specific time.

In order to obtain the eye classifier that takes the driver characteristics into account, the driver's identity information must first be verified to obtain an eye classifier that corresponds to the driver's identity. After the identity verification is completed, online recognition of driver fatigue state based on PERCLOS will be performed. First in the driving process, we use the ordinary car camera to obtain the driver's face image in real time. Next, the improved YOLOv3-tiny network is used to detect the driver's face. If the driver's face is detected, the facial area is used as the input image and the facial feature points are located using the Dlib toolkit. To reduce the false detection rate, the paper supplements the driver's head posture information as an auxiliary discrimination parameter. When the improved YOLOv3-tiny network fails to detect a face or locate the facial feature points, it is determined that the driver is in an abnormal head posture during driving, and this frame image is used as a closed-eye frame. After completing the face feature point positioning, the EFV is calculated based on the coordinates of the eye feature points, and then the driver's eye state classifier obtained by identity verification is used to determine the driver's eye state in the image. Finally, the number of closed-eye images of the driver is counted in a specific number of frames (1000 frames are set in the article), and then we calculate the PERCLOS value. If $PERCLOS > Th_{PERCLOS}$ ($Th_{PERCLOS}$ is the driver's fatigue state determination threshold (the article takes 0.4)), it is determined that the driver is in fatigue, otherwise, it is in non-fatigue.

2) FATIGUE JUDGMENT BASED ON BLINK FREQUENCY

Under normal circumstances, during the driving process, the driver blinks relatively quickly each time, and the duration is between 100-400ms. However, in the fatigue state, the duration of blinking is longer and more than 1 second, and the blinking frequency increases. Therefore, the blink frequency can also intuitively reflect the driver's fatigue level. The article stipulates that the system detects the state change of eye open-closed-open in turn as a blink, so the formula of blink frequency is:

$$F_{Blink} = \frac{N_{Blink}}{T} \times 100\% \tag{8}$$

where T is time, the unit is minute, and N_{Blink} is the number of blinks in T minutes.

Normal people blink about 10-20 times per minute when they are awake. When they are in fatigue, the blink frequency will increase by 64%. Based on related research, similar to section 2.5.1, after the identity verification is completed, online recognition of driver fatigue state based on blink frequency will be performed. Different from section 2.5.1, the number of blink images of the driver is counted in a specific number of frames (1000 frames are set in the article), and then we calculate the blink frequency. If $F_{Blink} > Th_{Blink}$ (Th_{Blink} is the driver's fatigue state determination threshold (the article takes 20)), it is determined that the driver is in fatigue, otherwise, it is in non-fatigue.

TABLE 1. Hardware configuration table.

Type	Specific parameters
Processor	Intel(R)Core(TM)i5-8400 CPU@2.80GHz 2.81GHz
GPU	NVIDIA GeForce GTX 1060 6GB
computer version	Windows 10
RAM	8.00GB
PythonTM version	3.6
Opencv version	3.4.2
Paddle version	1.5
CUDA version	9.2
CUDNN version	7.2

3) FATIGUE JUDGMENT BASED ON YAWN FREQUENCY

Similar to PERCLOS and blink frequency, yawn frequency is also an important indicator of evaluating fatigue. Yawn is a deep breathing activity that often occurs during laziness, tiredness, and lack of rest. And inhale more oxygen through enlarging the lung. It stimulates the central nervous system to boost the spirit, which is the conditioned reflex under fatigue. Using this conditioned reflex activity can provide an intuitive evaluation index of the fatigue level. In order to distinguish non-yawning mouth activities such as speaking, this article judges whether to yawn from the degree of mouth opening and the time of opening. The article stipulates that when the system detects the state changes of Close-Small-Big-Small-Close and the duration of opening up is more than 2 seconds, it is a yawn. The formula of yawn frequency is:

$$F_{Yawn} = \frac{N_{Yawn}}{T} \times 100\% \quad (9)$$

where T is time, the unit is minute, and N_{Yawn} is the number of blinks in T minutes.

According to related research, yawn frequency increases significantly when the human is in fatigue. Based on this, the paper studies driver fatigue based on yawn frequency. In order to obtain the mouth classifier that takes the driver characteristics into account, the driver's identity information must first be verified to obtain a mouth classifier that corresponds to the driver's identity. After the identity verification is completed, online recognition of driver fatigue state based on yawn frequency will be performed. First in the driving process, we use the ordinary car camera to obtain the driver's face image in real time. Next, the improved YOLOv3-tiny network is used to detect the driver's face. After completing the face feature point positioning, the MFV is calculated based on the coordinates of the mouth feature points, and then the driver's mouth state classifier obtained by identity verification is used to determine the driver's mouth state in the image. Finally, the number of yawn of the driver is counted in a specific number of frames (1000 frames are set in the article), and then we calculate the yawn frequency value. If $F_{Yawn} > Th_{Yawn}$ (Th_{Yawn} is the driver's fatigue state

determination threshold (the article takes 3)), it is determined that the driver is in fatigue, otherwise, it is in non-fatigue.

In summary, the flowchart of driver fatigue assessment model based on multi-feature fusion is shown in Figure 9.

III. EXPERIMENTS

To verify the validity of the algorithm, the paper evaluated the performance of the improved YOLOv3-tiny network with the Self-built data set DSD and public data set WIDER FACE. On this basis, the design comparison experiment is carried out to verify whether the fatigue driving detection algorithm based on facial multi-feature fusion is correct.

A. EXPERIMENTAL ENVIRONMENT AND DATA SET

The experimental platform is the Intel Core i5-8400 with x86 architecture, and the CPU clock speed is 2.80GHz. Graphics card is GTX1060 with Pascal architecture (CUDA: 9.2; CUDNN: 7.2), The RAM is 8G DDR4, and the opencv3.4.6 image library is used. The deep learning computing framework is PaddlePaddle1.5. The environment of the program is in python 3.6. Hardware configuration as shown in Table 1.

The data set used in the experiment included Self-built data set DSD and public data set WIDER FACE, where the public data set WIDER FACE includes 32203 pictures and 393703 marked faces, which is used to train Yolov3-tiny's face network. However, the WIDER FACE dataset only contains marker face images and does not provide any information about the driver's fatigue status. Therefore, the WIDER FACE data set cannot be used to analyze driver fatigue status. To this end, the driving state dataset (DSD) is established in this paper, which contains data collected by 50 test drivers sitting on a driving simulator (as shown in Figure 10). The data set of each test driver is shown in the table below.

B. FACE DETECTION

The improved YOLOv3-tiny network provides face landmarks for fatigue driving detection, and its performance is directly related to the pros and cons of the fatigue driving detection algorithm. Therefore, we quantitatively evaluate of
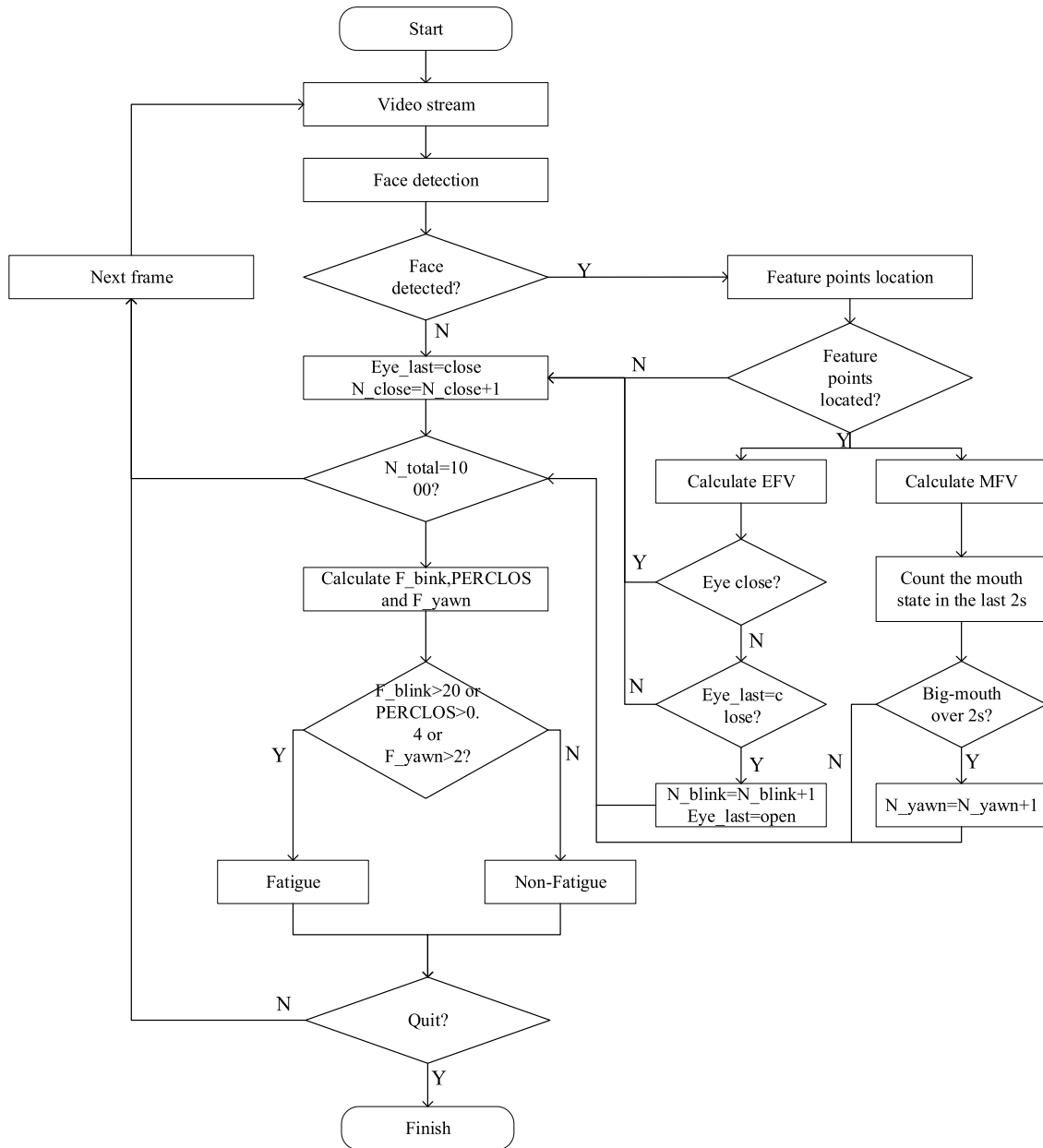
FIGURE 9. Flow chart of driver fatigue assessment model.

the performance of the improved YOLOv3-tiny network on the WIDER FACE data sets.

In this paper, accuracy are selected as evaluation indicators. It is an intuitive evaluation index of model performance., as shown in Equation(10).

$$accuracy = \frac{N_d}{N_t} \qquad (10)$$

where N_d is the number of correctly detected images, and N_t is the total number of images.

In the process of improving the YOLOv3-tiny network training and verification, the intersection ratio parameter (IOU) [57] is introduced to measure the similarity between

the face detection area and the marked real area. In Figure 11, *face_d* is the face area detected by the model, *face* is the real area marked, and the calculation formula is the Equation 11:

$$IoU = \frac{S(face\_d \cap face)}{S(face\_d \cup face)} \qquad (11)$$

where $S(face\_d \cap face)$ is the area of *face_d* $\cap$ *face*, and $S(face\_d \cup face)$ is the area of *face_d* $\cup$ *face*.

The intersection ratio indicates the degree of overlap between the model prediction area and the real area. As can be seen from Figure 11, the higher the value is, the higher the detection accuracy is. In the case IOU = 1, the prediction box overlaps with the real box. Generally speaking, in the task of

Data set	Normal driving videos	Fatigue driving video	Open eyes images	Closed eyes images	Big mouth images	Small mouth images	Closed mouth images
DSD	10	10	1000	3000	1000	1000	2000

FIGURE 10. Driving simulator for experiment.
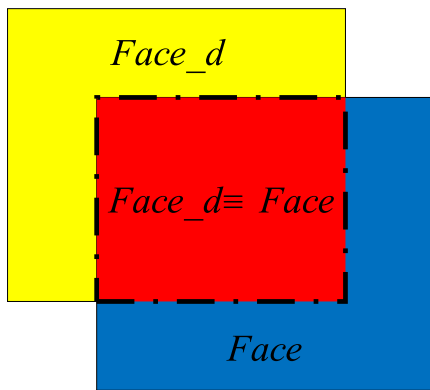
FIGURE 11. Intersection over union.

FIGURE 12. Driver face detection accuracy.

the target detection, it is considered that when the IOU > 0.5, the object is correctly detected. In the face detection task of this paper, considering that the face detection result directly affects the accuracy of the subsequent algorithm, we set a higher threshold. When the IOU>0.75, the face is considered to be correctly detected. Figure 12 shows the accuracy curve of the driver's face detection during the training of the improved YOLOv3-tiny network. Obviously, with the increase of training rounds, the accuracy of face detection gradually increases. The improved YOLOv3-tiny network has an accuracy rate of 97.9%.

C. FATIGUE STATE EVALUATION
1) ACCURACY
We use the DSD dataset to test the performance of fatigue detection. The DSD data set is shown in Table 2. Before the experiment, each subject needs to complete the identity entry. After this, the state classifiers required for online detection
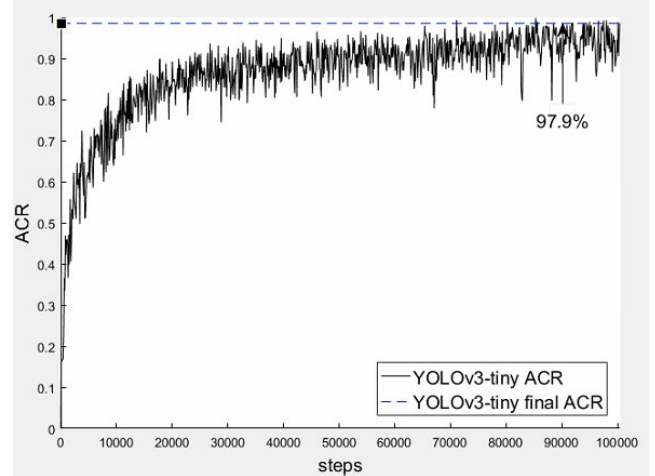
can be obtained. Then according to the state classifier of the corresponding subject, the eye state and mouth state of each frame in the video can be discriminated. Finally, calculate the PERCLOS, blink frequency, and yawn frequency of the corresponding subjects in this video respectively. If any of the indicators exceeds the set threshold, fatigue will be determined. We randomly select 10 videos from the data set, and the experimental results are shown in the table:

In this paper, we randomly select ten videos from the DSD test set, including non-fatigued driving status and fatigued driving status. The fatigue is judged based on the PERCLOS exceeding 0.4, the blink frequency exceeding 20, or the yawn frequency exceeding 3. As seen from the table, the accuracy of fatigue driving detection in 10 videos is 90%. In the end, the accuracy rate of the system was 95.10% accurate on the entire DSD data set. The system was tested to meet the expected design goals and meet the needs of practical applications.

2) SPEED
Based on hardware configuration as shown in Table 1, a comparison test is performed on the image source to verify the real-time performance of the system. The results are shown in Table 4.

From Table 4, the time taken by the image from the camera is slightly longer than by video. After analysis, it is considered that the image acquisition module based on OpenCV used different reading methods between the camera and the file video stream, so the time is different.

TABLE 3. Sample test table.

Sample number	PERCLOS	blink frequency	yawn frequency	Actual driving state	Predicted driving status
1	0.35	22.4	0	fatigued	fatigued
2	0.51	15.6	1.5	fatigued	fatigued
3	0.29	18.4	3.0	fatigued	fatigued
4	0.21	14.2	3.0	normal	fatigued
5	0.48	24.0	1.5	fatigued	fatigued
6	0.24	10.5	0	normal	normal
7	0.34	21.5	4.5	fatigued	fatigued
8	0.58	18.2	6.0	fatigued	fatigued
9	0.21	11.8	0	normal	normal
10	0.50	24.3	4.5	fatigued	fatigued

TABLE 4. The time spend of fatigue status judge.

Image source	Face Detection time	Facial feature point positioning time	Eye state acquisition time	Mouth state acquisition time	Total time
Camera	34.52ms	13.91ms	1ms	1ms	50.43ms
Video	32.13ms	13.91ms	1ms	1ms	48.04ms

TABLE 5. The time spend of fatigue status judge.

Algorithms	Accuracy/ %	speed/ $ms \cdot f^{-1}$
Adaboost + CNN	92.10	58.61
MTCNN+LRCN	91.00	52.97
Algorithm in this paper	94.32	49.43

Our algorithm shows that the system has good accuracy and high-speed performance under various conditions, and can accurately judge the fatigue state of the driver. Compared with Adaboost +CNN and MTCNN+LRCN algorithm [58], [59], our method improves the accuracy of the fatigue driving detection algorithm. It also has better real-time performance, which meets the requirements of the fatigue driving detection system. The comparative result is shown in Table 5.

IV. CONCLUSION

Fatigue driving can seriously affect driving skills and seriously threaten drivers and other traffic participants. At present, fatigue driving detection have achieved better research results, but it still needs to be improved, such as high intrusiveness, poor detection performance in complex environments, and simple evaluation indicator. Therefore, we propose a new detection algorithm for fatigue driving based on facial multi-feature fusion. The main contributions cover as follows.

◇ We designed a driver's face detection architecture based on the improved YOLOv3-tiny convolutional neural network, and trained the network with the open-source dataset WIDER FACE [28].

◇ We designed the eye and mouth SVM classifier that takes driver characteristics into account driver characteristics, which judges fatigue based on the actual driver's eyes and mouth size. It has high accuracy.

◇ We constructed the driver identity information library. There are three types of driver identity information in the system: driver biometrics, driver eye state classifier, and driver mouth state classifier. We train classifiers in advance and store them into the driver identity information library. Then, through identity verification, the driver's classifiers are called before system startup. It not only simplifies the initialization, but also avoids inaccuracies due to entering the identity manually.

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. 2018, doi: 10.1109/TITS.2018.2871262.

[2] Z. Zhou, Y. Zhou, Z. Pu, and Y. Xu, "Simulation of pedestrian behavior during the flashing green signal using a modified social force model," *Transportmetrica A, Transp. Sci.*, vol. 15, no. 2, pp. 1019–1040, Nov. 2019.

[3] Z. Zhou, Y. Cai, R. Ke, and J. Yang, "A collision avoidance model for two-pedestrian groups: Considering random avoidance patterns," *Phys. A, Stat. Mech. Appl.*, vol. 475, pp. 142–154, Jun. 2017.

[4] T. Toroyan, "Global status report on road safety 2013: Supporting a decade of action," *Injury Prevention*, vol. 15, no. 4, p. 286, 2013.

[5] D. Ma, X. Luo, S. Jin, D. Wang, W. Guo, and F. Wang, "Lane-based saturation degree estimation for signalized intersections using travel time data," *IEEE Intell. Transp. Syst. Mag.*, vol. 9, no. 3, pp. 136–148, Jul. 2017.

[6] D. Ma, X. Luo, W. Li, S. Jin, W. Guo, and D. Wang, "Traffic demand estimation for lane groups at signal-controlled intersections using travel times from video-imaging detectors," *IET Intell. Transp. Syst.*, vol. 11, no. 4, pp. 222–229, May 2017.

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

[8] 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.

[9] X. Qu, Y. Yu, M. Zhou, C.-T. Lin, and X. Wang, "Jointly dampening traffic oscillations and improving energy consumption with electric, connected and automated vehicles: A reinforcement learning based approach," *Appl. Energy*, vol. 257, Jan. 2020, Art. no. 114030.

[10] M. Zhou, Y. Yu, and X. Qu, "Development of an efficient driving strategy for connected and automated vehicles at signalized intersections: A reinforcement learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 1, pp. 433–443, Jan. 2020.

[11] Y. Feng and L. Zhaode, "Research on driver fatigue monitoring algorithm based on deep learning," *J. Wuhan Univ. Technol., Transp. Sci. Eng.*, vol. 42, no. 3, pp. 417–421, 2018.

[12] 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.

[13] D. Ma, X. Luo, S. Jin, W. Guo, and D. Wang, "Estimating maximum queue length for traffic lane groups using travel times from video-imaging data," *IEEE Intell. Transp. Syst. Mag.*, vol. 10, no. 3, pp. 123–134, Jun. 2018.

[14] C. Xu, Y. Yang, S. Jin, Z. Qu, and L. Hou, "Potential risk and its influencing factors for separated bicycle paths," *Accident Anal. Prevention*, vol. 87, pp. 59–67, Feb. 2016, doi: 10.1016/j.aap.2015.11.014.

[15] 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, Oct. 2018, doi: 10.1007/s11071-018-4421-9.

[16] 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.

[17] Z. Ning, R. Y. K. Kwok, K. Zhang, X. Wang, M. S. Obaidat, L. Guo, X. Hu, B. Hu, Y. Guo, and B. Sadoun, "Joint computing and caching in 5G-envisioned Internet of vehicles: A deep reinforcement learning-based traffic control system," *IEEE Trans. Intell. Transp. Syst.*, early access, Feb. 5, 2020, doi: 10.1109/TITS.2020.2970276.

[18] A. Anund, C. Fors, and C. Ahlstrom, "The severity of driver fatigue in terms of line crossing: A pilot study comparing day- and night time driving in simulator," *Eur. Transp. Res. Rev.*, vol. 9, no. 2, pp. 1–7, Jun. 2017, doi: 10.1007/s12544-017-0248-6.

[19] 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.

[20] Z. Ning, P. Dong, X. Wang, X. Hu, L. Guo, B. Hu, Y. Guo, T. Qiu, and R. Y. Kwok, "Mobile edge computing enabled 5G health monitoring for Internet of medical things: A decentralized game theoretic approach," *IEEE J. Sel. Areas Commun.*, pp. 1–16, Mar. 2020. [Online]. Available: https://www.researchgate.net/publication/339874488

[21] T. Abe, T. Nonomura, Y. Komada, S. Asaoka, T. Sasai, A. Ueno, and Y. Inoue, "Detecting deteriorated vigilance using percentage of eyelid closure time during behavioral maintenance of wakefulness tests," *Int. J. Psychophysiol.*, vol. 82, no. 3, pp. 269–274, 2011.

[22] M. S. Devi and P. R. Bajaj, "Fuzzy based driver fatigue detection," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Oct. 2010, pp. 3139–3144.

[23] H. Singh, J. S. Bhatia, and J. Kaur, "Eye tracking based driver fatigue monitoring and warning system," in *Proc. India Int. Conf. Power Electron. (IICPE)*, Jan. 2011, pp. 1–6.

[24] A. Haro, M. Flickner, and I. Essa, "Detecting and tracking eyes by using their physiological properties, dynamics, and appearance," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2000, pp. 163–168.

[25] R. C. Coetzer and G. P. Hancke, "Eye detection for a real-time vehicle driver fatigue monitoring system," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2011, pp. 66–71.

[26] Z. Ning, Y. Li, P. Dong, X. Wang, M. S. Obaidat, X. Hu, L. Guo, Y. Guo, J. Huang, and B. Hu, "When deep reinforcement learning meets 5G-enabled vehicular networks: A distributed offloading framework for traffic big data," *IEEE Trans. Ind. Informat.*, vol. 16, no. 2, pp. 1352–1361, Feb. 2020.

[27] 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.

[28] S. Yang, P. Luo, C.-C. Loy, and X. Tang, "WIDER FACE: A face detection benchmark," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, Dec. 2016, pp. 5525–5533.

[29] J.-H. Won, D.-H. Lee, K.-M. Lee, and C.-H. Lin, "An improved YOLOv3-based neural network for de-identification technology," in *Proc. 34th Int. Tech. Conf. Circuits/Syst., Comput. Commun. (ITC-CSCC)*, JeJu, South Korea, Jun. 2019, pp. 1–2.

[30] X. Chen, X. Luo, X. Liu, and J. Fang, "Eyes localization algorithm based on prior MTCNN face detection," in *Proc. IEEE 8th Joint Int. Inf. Technol. Artif. Intell. Conf. (ITAIC)*, Chongqing, China, May 2019, pp. 1763–1767.

[31] Z. Ning, J. Huang, X. Wang, J. J. P. C. 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. 2019.

[32] 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 crowd-sourced data," *IEEE Trans. Ind. Informat.*, vol. 16, no. 2, pp. 1436–1444, Feb. 2020.

[33] P. Viola and M. Jones, "Robust real-time face detection," in *Proc. 8th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 2. Vancouver, BC, Canada: Institute of Electrical and Electronics Engineers, 2001, p. 747.

[34] S. Saha and V. Demoulin, "ALOHA: An efficient binary descriptor based on Haar features," in *Proc. 19th IEEE Int. Conf. Image Process.*, Orlando, FL, USA, Sep. 2012, pp. 2345–2348.

[35] S. Jida, B. Aksasse, and M. Ouanan, "Face segmentation and detection using Voronoi diagram and 2D histogram," in *Proc. Intell. Syst. Comput. Vis. (ISCV)*, Fez, Morocco, Apr. 2017, pp. 1–5.

[36] J. Zou and R. Song, "Microarray camera image segmentation with faster-RCNN," in *Proc. IEEE Int. Conf. Appl. Syst. Invention (ICASI)*, Chiba, Japan, Apr. 2018, pp. 86–89.

[37] N. L. Arifin, H. Widiastuti, and A. Wibowo, "Study on effect of source to film distance (SFD) on the radiographic images," in *Proc. Int. Conf. Appl. Eng. (ICAE)*, Batam, Indonesia, Oct. 2018, pp. 1–4.

[38] P. Samangouei, R. Chellappa, M. Najibi, and L. S. Davis, "Face-MagNet: Magnifying feature maps to detect small faces," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Lake Tahoe, NV, USA, Mar. 2018, pp. 122–130.

[39] M. El-Arabawy, S. Zaki, and F. Harby, "Improved AdaBoost algorithm for face detection," in *Proc. Int. Conf. Image Process., Comput. Vis., Pattern Recognit.*, Las Vegas, NV, USA, vol. 1, 2010, pp. 353–358.

[40] D. P. Lestari, R. Kosasih, T. Handhika, Murni, I. Sari, and A. Fahrurozi, "Fire hotspots detection system on CCTV videos using you only look once (YOLO) method and tiny YOLO model for high buildings evacuation," in *Proc. 2nd Int. Conf. Comput. Informat. Eng. (ICIE)*, Banyuwangi, Indonesia, Sep. 2019, pp. 87–92.

[41] D. Goularas and S. Kamis, "Evaluation of deep learning techniques in sentiment analysis from Twitter data," in *Proc. Int. Conf. Deep Learn. Mach. Learn. Emerg. Appl. (Deep-ML)*, İstanbul, Turkey, Aug. 2019, pp. 12–17.

[42] S. Kido, Y. Hirano, and N. Hashimoto, "Detection and classification of lung abnormalities by use of convolutional neural network (CNN) and regions with CNN features (R-CNN)," in *Proc. Int. Workshop Adv. Image Technol. (IWAIT)*, Chiang Mai, Thailand, Jan. 2018, pp. 1–4.

[43] Y. Tian, Y. Du, Q. Zhang, J. Cheng, and Z. Yang, "Depth estimation for advancing intelligent transport systems based on self-improving pyramid stereo network," *IET Intell. Transp. Syst.*, vol. 14, no. 5, pp. 338–345, May 2020.

[44] Y. Tian, Q. Zhang, Z. Ren, F. Wu, P. Hao, and J. Hu, "Multi-scale dilated convolution network based depth estimation in intelligent transportation systems," *IEEE Access*, vol. 7, pp. 185179–185188, 2019.

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

[46] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, pp. 1189–1232, Oct. 2001.

[47] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[48] Q. Zhang and S.-I. Kamata, "Improved color barycenter model and its separation for road sign detection," *IEICE Trans. Inf. Syst.*, vol. E96.D, no. 12, pp. 2839–2849, Dec. 2013.

[49] X. Cao, Y. Wei, F. Wen, and J. Sun, "Face alignment by explicit shape regression," *Int. J. Comput. Vis.*, vol. 107, no. 2, pp. 177–190, Apr. 2014.

[50] P. Dollár, P. Welinder, and P. Perona, "Cascaded pose regression," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, San Francisco, CA, USA, Jun. 2010, pp. 1078–1085.

[51] Q. Zhang and S.-I. Kamata, "A novel color descriptor for road-sign detection," *IEICE Trans. Fundam. Electron., Commun. Comput. Sci.*, vol. E96-A, no. 5, pp. 971–979, May 2013.

[52] 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.

[53] D. Ma, X. Song, and P. Li, "Daily traffic flow forecasting through a contextual convolutional recurrent neural network modeling inter-and intra-day traffic patterns," *IEEE Trans. Intell. Transp. Syst.*, early access, Feb. 24, 2020, doi: 10.1109/TITS.2020.2973279.

[54] D. Ma, J. Xiao, X. Song, X. Ma, and S. Jin, "A back-pressure-based model with fixed phase sequences for traffic signal optimization under oversaturated networks," *IEEE Trans. Intell. Transp. Syst.*, early access, Apr. 29, 2020, doi: 10.1109/TITS.2020.2987917.

[55] Z. You, Y. Gao, J. Zhang, H. Zhang, M. Zhou, and C. Wu, "A study on driver fatigue recognition based on SVM method," in *Proc. 4th Int. Conf. Transp. Inf. Saf. (ICTIS)*. Banff, AB, Canada: Institute of Electrical and Electronics Engineers, Aug. 2017, pp. 693–697.

[56] J.-J. Yan, H.-H. Kuo, Y.-F. Lin, and T.-L. Liao, "Real-time driver drowsiness detection system based on PERCLOS and grayscale image processing," in *Proc. Int. Symp. Comput., Consum. Control (ISC)*, Xi'an, China, Jul. 2016, pp. 243–246.

[57] L. Tychsen-Smith and L. Petersson, "Improving object localization with fitness NMS and bounded IoU loss," in *Proc. 31st Meeting IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Salt Lake City, UT, USA, Jun. 2018, pp. 6877–6885.

[58] G. Lei, X. Liang, Z. Xiao, and Y. Li, "Real-time driver fatigue detection based on morphology infrared features and deep learning," *Infr. Laser Eng.*, vol. 47, no. 2, 2018, Art. no. 203009.

[59] L. Tychsen-Smith and L. Petersson, "Improving object localization with fitness NMS and bounded IoU loss," 2017, *arXiv:1711.00164*. [Online]. Available: https://arxiv.org/abs/1711.00164

KENING LI received the B.E. degree in communications and transportation from the University of Henan Agricultural University, Henan, China, in 2004, the M.S. degree in vehicle operation engineering from Jilin University, Jilin, in 2007, and the Ph.D. degree in vehicle engineering from Shanghai Jiao Tong University, Shanghai, in 2014. He is currently a Lecturer with the School of Traffic and Environment, Shenzhen Institute of Information Technology, Shenzhen, China. His research interests include vehicle system dynamics and control, intelligent vehicle systems, and artificial intelligence.

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 research interests include intelligent vehicles, computer vision, and 3D laser radar.
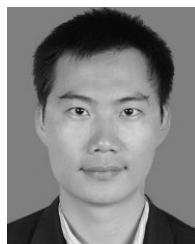
ZILIANG REN (Member, IEEE) received the Ph.D. degree from the South China University of Technology, Guangzhou, China, in 2017. He is currently with the Guangdong Provincial Key Laboratory of Robotics and Intelligent System, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, as a Postdoctoral Researcher and an Assistant Researcher. His current research interests include computer vision and machine learning.

● ● ●