

Fatigue Driving Detection Based on Deep Learning and Multi-Index Fusion

HUIJIE JIA¹, ZHONGJUN XIAO², AND PENG JI³

School of electrical engineering and automation, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353, China

Corresponding author: Zhongjun Xiao (jinanaxiaoxiao2012@163.com).

This work was supported in part by the National Natural Science Foundation of China (research on the theory and method of visual gesture remote operation control of ground reconnaissance robot in all-day complex background) under Grant 61903207.

ABSTRACT In order to reduce traffic accidents caused by fatigue driving, a fatigue driving detection algorithm is proposed based on deep learning and facial multi-index fusion from the driver's facial features. Because the scene in the actual driving process is very complex and changeable, this algorithm first improves the multi-task cascaded convolutional neural network (MTCNN) so that it can quickly and accurately locate the face and detect the facial key points. According to the facial key points, the driver's eyes and mouth regions are determined. Second, these regions are input into the eyes and mouth state recognition network (E-MSR Net) for state recognition. The E-MSR Net is a depth separable convolution neural network that is improved and optimized based on MobilenetV2. Finally, the three facial features of eye closure rate (ECR), mouth opening rate (MOR), and head non-positive face rate (HNFR) are fused to judge the driver's fatigue state. This algorithm can quickly and accurately make judgments in the face of complex and changeable scenes. At the same time, it can avoid the failure of the algorithm caused by the occlusion of the eyes or mouth due to wearing sunglasses or masks during driving. The accuracy of the proposed algorithm on the self-made data set achieved 97.5%, which proved the feasibility of the algorithm.

INDEX TERMS Fatigue driving detection, improved MTCNN, E-MSR Net, facial multi-index fusion.

I. INTRODUCTION

Fatigue driving endangers road traffic safety and has grown up to be an important cause of traffic accidents. A 2011 World Health Organization survey found that 1.3 million people died each year from road traffic accidents and about 50 million were disabled. By 2020, traffic accidents were the fifth leading cause of death worldwide, killing about 2.4 million people each year. The American Automobile Association (AAA) Traffic Safety Foundation showed that nearly one-fifth of the deaths in traffic accidents in the United States each year, about 4.55 million people were killed, which was caused by fatigue driving.

How to effectively monitor and determine the fatigue driving state so as to achieve fatigue warning has become a hot topic in scientific research. At present, there are three main driving fatigue detection methods. The first is the driver fatigue detection technology based on driving data [1] [2]. This technology detects the driver's fatigue state by monitoring the real-time driving trajectory of the vehicle, the rotational speed of the steering wheel under a certain hand-held pressure [3], and the driving monitoring device to monitor the lane line deviation degree and other indirect driving data in real-time. Fatigue detection based on driving

data is non-invasive and does not affect the normal driving behavior during monitoring. When the monitoring shows that the driving trajectory of the vehicle is deviated, or the pressure of the hand-held steering wheel is reduced. The system will determine that the driver is currently fatigued, and the driver will be awakened by vibration and voice. The indirect driving data of the vehicles will also be affected by the driver's personal driving habits, driving skills and other non-fatigue factors.

Wang *et al.* [4] combined with the steering wheel angle information at different time points, the longitudinal acceleration and lateral acceleration of the vehicle, and used the random forest algorithm to predict the fatigue state of the driver. After comparing and analyzing the above indicators, they found that the lateral acceleration of the vehicle is the best indicator to detect the fatigue state of the driver. Mao *et al.* [5] used a simulated driver to obtain a variety of physical characteristics of the vehicle during driving, analyzed the normal state and abnormal state of the vehicle, and constructed a classifier according to different characteristics.

The second is the driving fatigue state detection method based on driver physiological parameters [6] [7] [8] [9]. The signals from electrode patches at the positions of forehead,

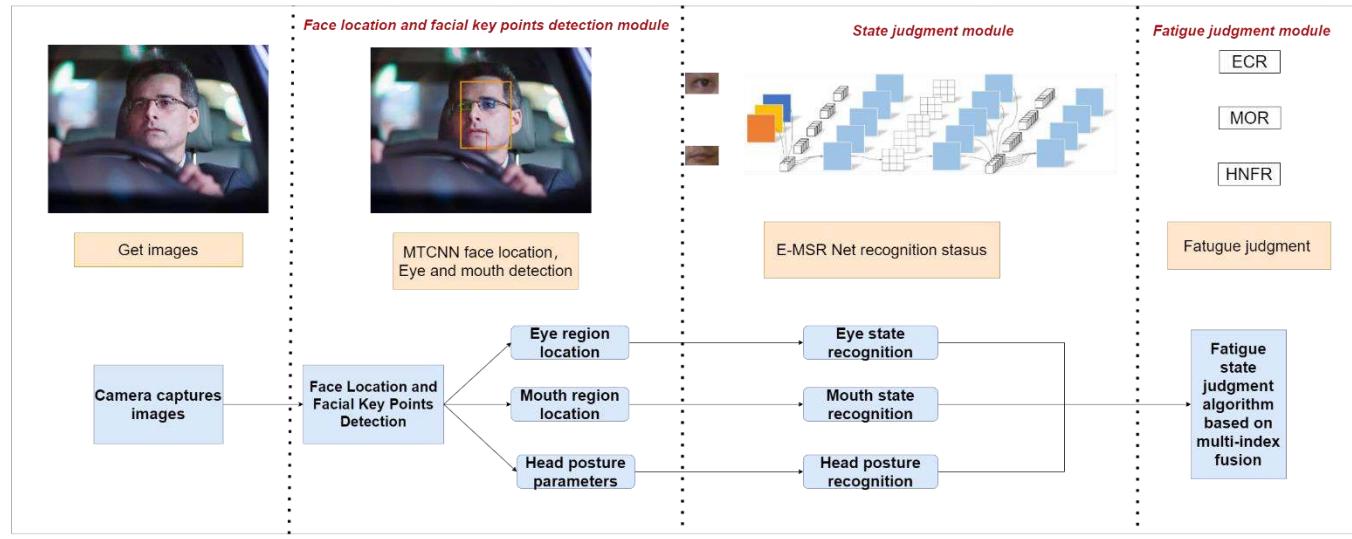


FIGURE 1. Fatigue detection algorithm framework

heart and muscle are received through special medical equipment, and the signals are transmitted to the system for real-time analysis [10] [11]. The clinically defined physiological parameter signal threshold is used as the basis for fatigue identification. [12] Physiological parameters are mostly collected by special signal acquisition instruments. In order to obtain the precision of the collected data and the correctness of the analysis, the volume of medical equipment is relatively large, and it will occupy the driving space. The driver cannot fully carry out the driving of the vehicle, which invisibly aggravates the occurrence of accidents. In addition, because one end of the data acquisition equipment must be connected to the skin of the human body, or placed on the head, or placed on the trunk, it will more or less will restrict the behavior of the driver, increasing the uncertainty of the driving process. Most importantly, due to the individual differences among drivers, the parameter thresholds of each driver's fatigue identification are not the same, which brings resistance to individual differences to fatigue detection. The system design needs to be more targeted and cannot be universal.

Chui et al. [13] proposed an EEG signal processing method based on SVM algorithm. This method reduces the amount of calculation. At the same time, it speeds up the operation speed, which is beneficial to the improvement of fatigue driving detection speed. Piazz et al. found that when the driver's heart beat frequency during fatigue is lower than 20% of the normal state, the driver appears fatigue phenomenon, which can be judged by the driver's heart beat frequency.

The third is the fatigue detection method based on computer vision [14] [16]. For fatigued drivers, some visual changes can be easily observed from their facial features, such as eye, head and face changes, including longer blinking time, slow eyelid movement, smaller degree of eye

opening (or even close the eyes), frequent nodding, yawning, gaze (narrow line of sight), slow facial expression and drooping posture [17]. Computer vision is a natural and non-invasive method. The facial visual features that can characterize the driver's fatigue level are extracted from the images captured by the camera in front of the driver. Combined with machine vision, image processing, pattern recognition and other related technologies, the driver's eye state, mouth state, and head motion state are analyzed to judge the driver's fatigue state [18] [19]. Compared with many traditional methods, the fatigue detection method based on computer vision has the advantages of non-contact, non-interference and high detection accuracy.

Zhang et al. [20] designed a new yawning detection method for fatigue detection. They used a convolution neural network to extract spatial image features and long-term and short-term memory network to analyze time features. This method can solve the problem of eye occlusion caused by sunglasses. However, under the condition of insufficient sunlight at night, the algorithm will miss detection and check errors. Zhuang et al. [21] proposed an effective fatigue detection method based on eye status with pupil and iris segmentation. The method separates the pupil and iris of the human eye through a streamlined network, and then determines the eye state according to the characteristics of the pupil and iris. This method is less affected by the change of sunlight, but it is not very effective for wearing sunglasses. Zhang et al. [22] proposed an algorithm using image processing technology to process the images, and then evaluated it by SVM model. Finally, the sequence floating forward selection algorithm is used to select the optimal parameters, and the fatigue detection model was established. Li et al. [34] used the improved Yolov3-micro to extract facial features, and then judged the driver's eyes and mouth. This method combines a variety of deep learning algorithms

to improve the accuracy of fatigue detection. However, the deep learning network model used is large, which is not conducive to porting mobile terminals.

The existing detection methods have the problems of low comfort, vulnerability to external factors such as sunlight, and low accuracy caused by a single index. To solve these problems, this paper studies a driving fatigue detection algorithm based on deep learning and facial multi-index fusion, which mainly includes face location and facial key points detection module, state judgment module, fatigue judgment module.

II. FATIGUE DETECTION ALGORITHM

This algorithm first collects the driver's images through the camera in front of the driver, and then the images use the improved MTCNN (Multi-task Convolutional Neural Network) to locate the face and detect the facial key points. Based on the detected key points, the location areas of the eyes and mouth are determined, and they are input into the eye and mouth state recognition network (E-MSR Net) for state judgment. Finally, the fatigue state of the driver is judged by the facial multi-index fusion strategy. The overall framework of the algorithm in this paper is shown in Figure 1.

A. FACE LOCATION AND FACIAL KEY POINTS DETECTION BASED ON IMPROVED MTCNN

MTCNN contains three cascaded multi-task convolutional neural networks, namely Proposal Network(P-Net), Refine Network(R-Net), and Output Network (O-Net) [23]. Each multi-task convolutional neural network has three learning tasks, namely face classification, border regression, and key point positioning. However, the focus of the three stages is difference.

The network model in the first stage is called the P-Net. The network structure is shown in Figure 2. The main function is to obtain the window of the face region and the boundary box regression. The obtained window of the face area will be corrected by the results of the boundary box regression, and then use non-maximum suppression (NMS) to merge the overlapping windows.

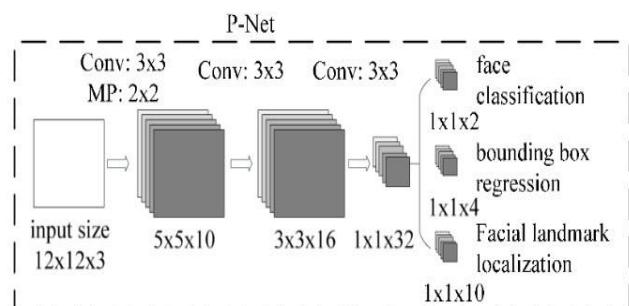


FIGURE 2. P-Net structure

The network model in the second stage is called the R-Net, which filters out most of the non-face candidate windows

through a more capable CNN network, then continues to correct the results of the Bounding Box regression, and uses NMS to merge overlapping windows. The network structure is shown in Figure 3

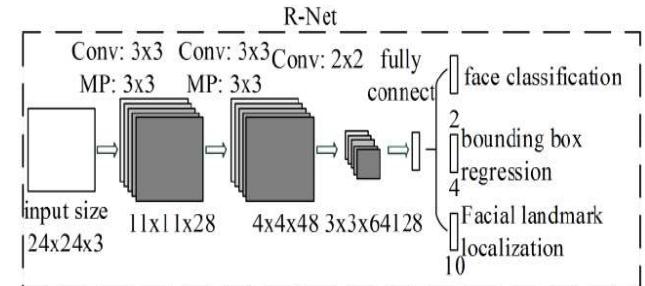


FIGURE 3. R-Net structure

The network model in the third stage is called the O-Net. Inputting the data of the second stage for further extraction, and find the five marker points on the face through a more capable network. The network structure is shown in Figure 4.

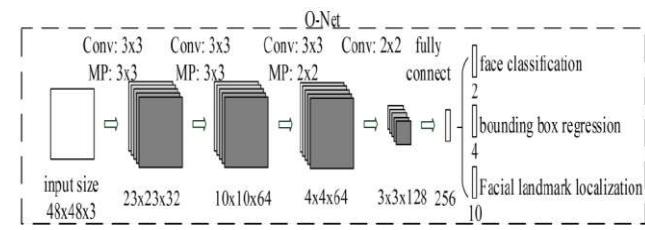
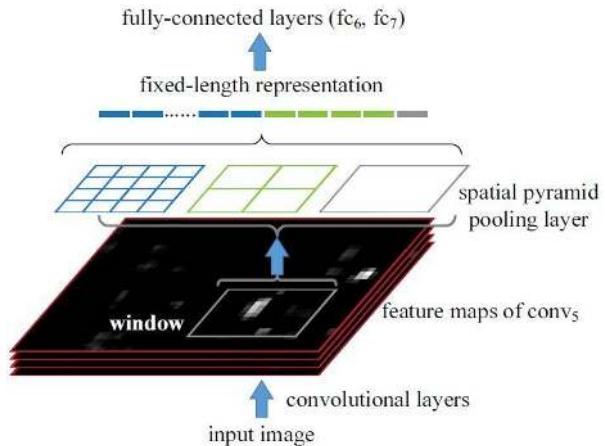


FIGURE 4. O-Net structure

In the process of fatigue detection, due to changes in human posture and scene light intensity, MTCNN often fails to detect. In response to this phenomenon, MTCNN is improved. Adding SPP layer to the last layer of O-Net network can improve the accuracy of the network in complex scenes. Adding BN algorithm to MTCNN can improve network performance.

1) ADDING THE SPATIAL PYRAMID POLLING(SPP) LAYER TO O-NET NETWORK STRUCTURE

Aiming at the problem of complex scene changes in the driving process, an SPP layer [24] is added after the last convolution layer in the O-Net network, which can output a fixed-length vector for feature maps of different sizes. As shown in Figure 5, the SPP layers of this algorithm adopt 1×1 , 2×2 , and 4×4 for the three-proportion pyramid pooling layers. After the SPP layer, the feature map is divided into 21 parts, and the maximum pooling operation is performed on each part respectively. Through the SPP layer, the feature map is transformed into 21×128 matrices, which is sent to the full connection layer to develop into a one-dimensional matrix. By adding the SPP layer, the feature maps of any size can be converted into feature vectors of fixed size, which improves the recognition accuracy of the network model for features of different scenes and sizes, and reduces the amount of calculation.

**FIGURE 5.** The SPP layer specific workflow

2) ADDING THE BATCH NORMALIZATION (BN) ALGORITHM BEFORE EACH LAYER OF MTCNN NETWORK ACTIVATION FUNCTION

The different distribution of input values will affect the training of MTCNN network. When the input feature values have relatively large differences, with the training of the network, these differences will affect the back-layer network, and at the same time will lead to gradient explosion in the back propagation process.

The function of the BN algorithm [25] is to standardize the input value and reduce the difference of input value to the same range. On the one hand, it improves the convergence degree of the gradient and accelerates the training speed of the model. On the other hand, each layer can face the input value of the same feature distribution as much as possible, which reduces the uncertainty caused by the changes, and also reduces the impact on the back layer network, making each layer network become relatively independent, alleviating the problem of gradient disappearance in training, so as to improve the performance of the network.

B. EYE AND MOUTH STATE RECOGNITION NETWORK (E-MSR NET)

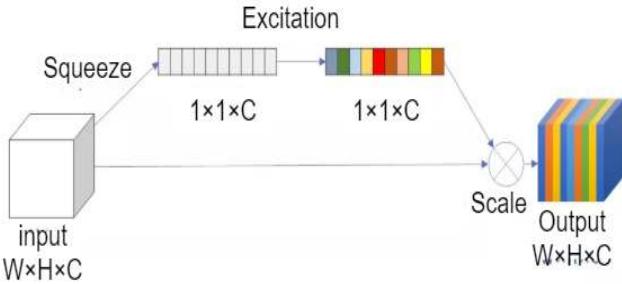
After these images pass through the face location and facial key points detection module, identifying the eye and mouth areas of the face through the detected feature points. Sending them to the depth separable convolution neural network E-MSR Net to judge the state of the driver's eyes and mouth. The E-MSR Net is a lightweight network, which adds SE network to the inverted residual structure of MobileNetV2 network [26] to improve the network accuracy, and replaces the sigmoid activation function in MobileNetV2 network with h-swish activation function to reduce the calculation amount.

1) ADDING THE SQUEEZE-AND-EXCITATION (SE) NETWORKS

The SE networks [27] mainly includes squeeze and excitation, as shown in Figure 6. Firstly, the global average pooling operation is performed on the input feature map. Next, two full connection layers are connected to increase the nonlinear

processing of the feature map and fit the complex correlation between channels.

By adding the SE network into the bottleneck of the MobileNetV2 network, as shown in Figure 7. The E-MSR Net can automatically obtain the importance of each feature channel through learning, and then according to this result to enhance useful features and inhibit features that are not useful for the current tasks. The feature is calibrated to make the effective weight is significant and the ineffective weight is small, so as to improve the accuracy of the network.

**FIGURE 6.** The SE Net composition

2) CHANGING THE ACTIVATION FUNCTION

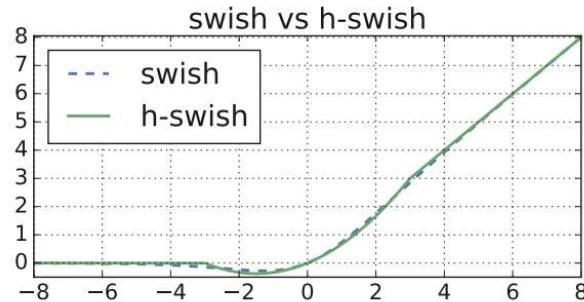
The sigmoid function [28] used in the MobileNetV2 network can be seen from the formula that the sigmoid activation function is very complex in the calculation and derivation process, which consumes a lot of computing resources. This algorithm uses the h-swish function to replace it. The formula is shown below.

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (1)$$

$$h\text{-swish}[x] = x \frac{\text{ReLU6}(x+3)}{6} \quad (2)$$

$$\text{ReLU} = \max(0, x) \quad (3)$$

The h-swish function has many advantages while maintaining accuracy. Firstly, ReLU6 can be implemented in many hardware and software frameworks. Secondly, it avoids the loss of numerical accuracy when quantifying. And because of the decrease in the amount of calculation, the running speed is accelerated. The function image is shown in Figure 8.

**FIGURE 8.** H-swish activation function diagram

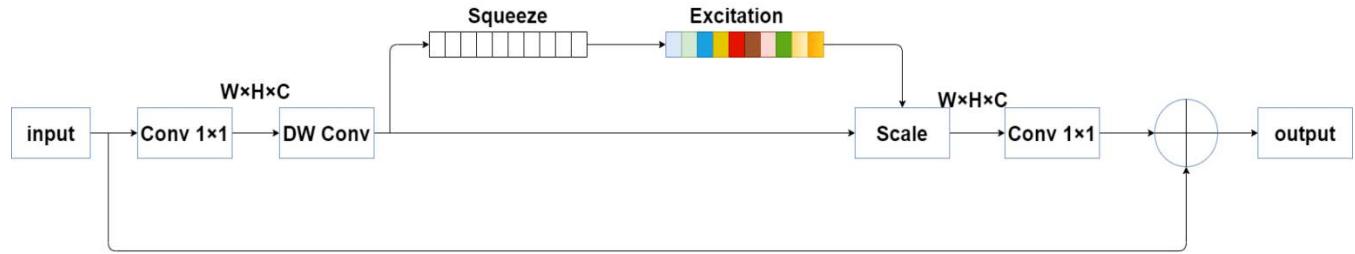


FIGURE 7. The bottleneck of the E-MSR Net

III. FATIGUE STATE JUDGMENT ALGORITHM BASED ON MULTI-INDEX FUSION

In order to solve the problem that the algorithm is inaccurate due to a single evaluation index in the process of fatigue detection, this algorithm defines three evaluation indexes of eye closure rate (ECR), mouth opening rate (MOR) and head non-face rate (HNFR). The three evaluation indexes are fused to determine whether the driver is tired or not.

A. EYE STATE EVALUATION INDEX

PERCLOS [29] is an internationally recognized fatigue judgment criterion, which refers to the time proportion of eyes closed within a certain time. The calculation formula is as follows.

$$\text{PERCLOS} = \frac{\text{Number of closed eyes frames}}{\text{Total frame number of detection period}} \times 100\% \quad (4)$$

The PERCLOS judgment criterion includes P70, P80, and EM, which respectively indicate that the area of eyelid covering the pupil exceeds 70 %, 80 % and 50 % is counted as eye closure. The time proportion of eye closure in a certain time is counted. Among them, P80 is considered to be the most responsive standard to fatigue. The principle is shown in Figure 9.

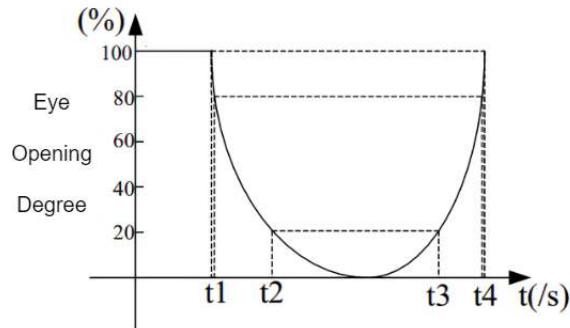


FIGURE 9. PERCLOS schematic diagram under P80 standard

In this paper, the eye state evaluation index refers to the P80 criterion in the PERCLOS criterion. Definition f_e the ECR to judge the driver's eye fatigue state. The calculation formula is as follows:

$$f_e = \frac{t_e}{T_e} \times 100\% \quad (5)$$

Among them, t_e represents the number of eye closing frames during the detection time, and T_e represents the total number of frames during the detection time.

B. MOUTH STATE EVALUATION INDEX

The mouth state evaluation index is similar to the eye state evaluation index. Definition f_m represents the MOR to judge the driver's mouth fatigue state. The calculation formula is as follows:

$$f_m = \frac{t_m}{T_m} \times 100\% \quad (6)$$

Among them, t_m represents the number of mouths opening frames during the detection time, and T_m represents the total number of frames in the detection time.

C. HEAD POSTURE EVALUATION INDEX

The improved MTCNN detects the five 2D key points of the left eye, right eye, nose tip, and left and right mouth corners of the human face. The POSIT algorithm converts five 2D feature points into five 3D feature points in the world coordinate systems by rotation, translation and other methods. Next, the transformation parameters are estimated. Finally, the human head posture parameters in 2D plane are obtained, which are yaw angle (Yaw), pitch angle (Pitch) and roll angle (Roll). According to the actual head posture when fatigued during driving, Pitch and Roll are selected as the evaluation indexes.

The head posture evaluation index is similar to the eye and mouth state evaluation index. Definition f_h represents the HNFR to judge the driver's head posture state. The calculation formula is as follows:

$$f_h = \frac{t_h}{T_h} \times 100\% \quad (7)$$

Among them, t_h represents the number of head posture fatigue frames during the detection time, and T_h represents the total number of frames in the detection time.

D. FACIAL MULTI-INDEX FUSION

In the actual driving environment, fatigued driving is a complex state of psychology and physiology, and the detection results are easily interfered with by various environments. If only eye features or mouth features are extracted to determine the driver's fatigue, the accuracy of detection will be greatly reduced when the driver wears sunglasses or masks and other facial occlusions. Therefore,

the fatigue detection method based on a single feature needs to be carried out in an ideal situation. This algorithm uses the fusion of ECR, MOR, HNFR three evaluation indicators to determine the driver's fatigue, in order to improve the system performance and enhance the system robustness.

According to the single eye closure time of many tests under different eye states, the eye state parameters were calculated [30]. In the normal state, the time of single eye closure is usually 0.12 ~ 0.15 seconds. In the fatigue state, the time of single eye closure is greater than or equal to 0.5 seconds, and the time is significantly greater than that in an awake state. Therefore, setting the parameter threshold of ECR as 0.5, that is, when the eye closure frequency within unit time is 0.5, it is judged that the eyes are in a state of fatigue. According to people usually yawn state duration is 3 ~ 5 seconds [31]. In this algorithm, the unit detection time is selected as 20 seconds, and the yawning frequency within 20 seconds cannot exceed twice. Therefore, this paper sets the parameter threshold of MOR as 0.3, that is, when the yawning frequency within unit time is 0.3, it is judged that the mouth is in a state of fatigue.

According to the PERCLOS judgment criterion, when the Pitch and Roll change angles of the head posture are detected more than 20 %, it is judged that the head is in an incorrect state. According to multiple tests, the head posture changes when the driver is fatigued during actual driving, the parameter threshold of HNFR is set to 0.5. That is, when the proportion of the driver's improper head posture in the unit time period is 0.5, the head posture is judged to be fatigue.

The parameter threshold fusion of the above ECR, MOR, and HNFR three evaluation indexes is used to determine whether the driver is tired. The specific evaluation indexes are as follows:

$$\begin{cases} \text{normal: } f_e < 0.5 \text{ and } f_h < 0.5 \text{ and } f_m < 0.3 \\ \text{danger: } f_e \geq 0.5 \text{ or } f_h \geq 0.5 \text{ or } f_m \geq 0.3 \end{cases} \quad (8)$$

IV. EXPERIMENT AND RESULT ANALYSIS

A. FACE LOCATION AND KEY POINTS DETECTION MODEL TRAINING AND RESULT ANALYSIS

In this algorithm, the improved MTCNN uses the WIDER FACE database [32] as the training data set of face location. The WIDER FACE database is the mainstream face detection database, which has 32203 face images and 393703 labeled faces. Moreover, all kinds of scenes are very complex, which can evaluate the performance of the algorithm from all directions and angles. Using the MTFL database as the training data for facial feature points. The MTFL database contains 12995 face images and 5 key points. It also provides information on gender, smile, glasses, and head posture. The images in the data set section are shown below. The dataset is shown in Figure 10

This experimental computer hardware configuration is shown in Table 1. The experiment in this paper is based on python3.7 and Tensorflow2.2 in the Windows10 environment.

CPU uses Intel (R) Core (TM) i7 - 10870H, and GPU uses NVIDIA GeForce RTX 2060 for model training. The experimental results are shown in Figure 11.

TABLE 1. Hardware configuration table

Type	Parameter
CPU	Intel(R) Core(TM) i7-10870H CPU
GPU	NVIDIA GeForce RTX 2060
CUDA version	CUDA 10.1
system environment	Windows 10

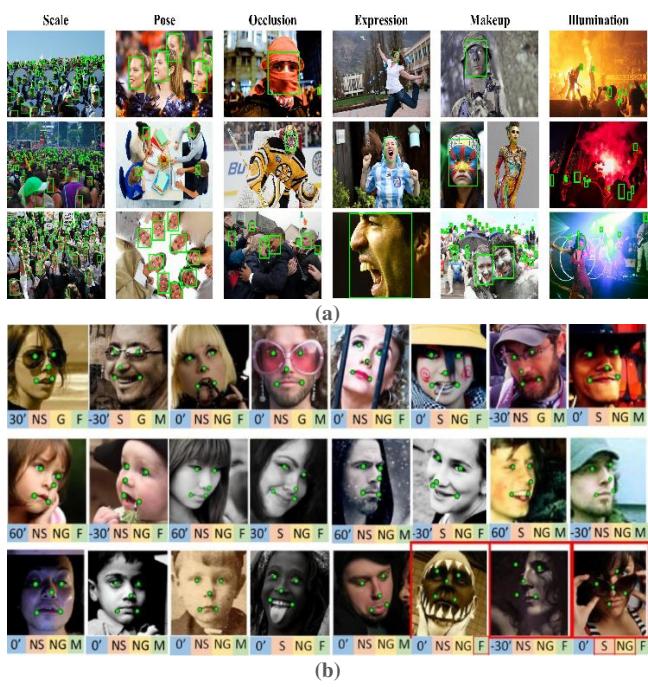


FIGURE 10. Face location and facial key points detection module dataset.
(a) WIDER FACE dataset; (b) MTFL dataset

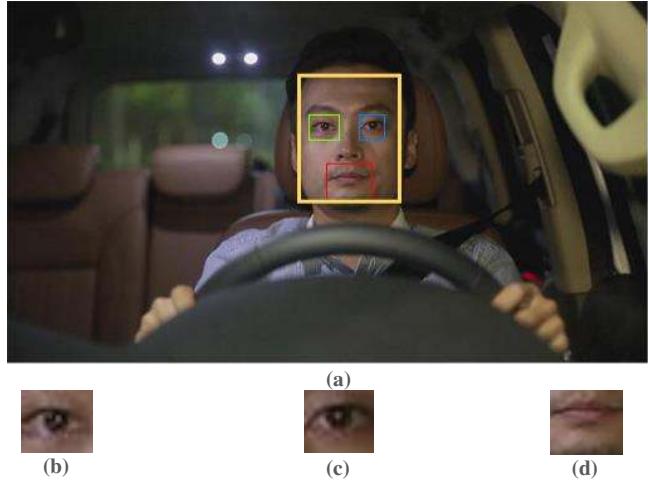
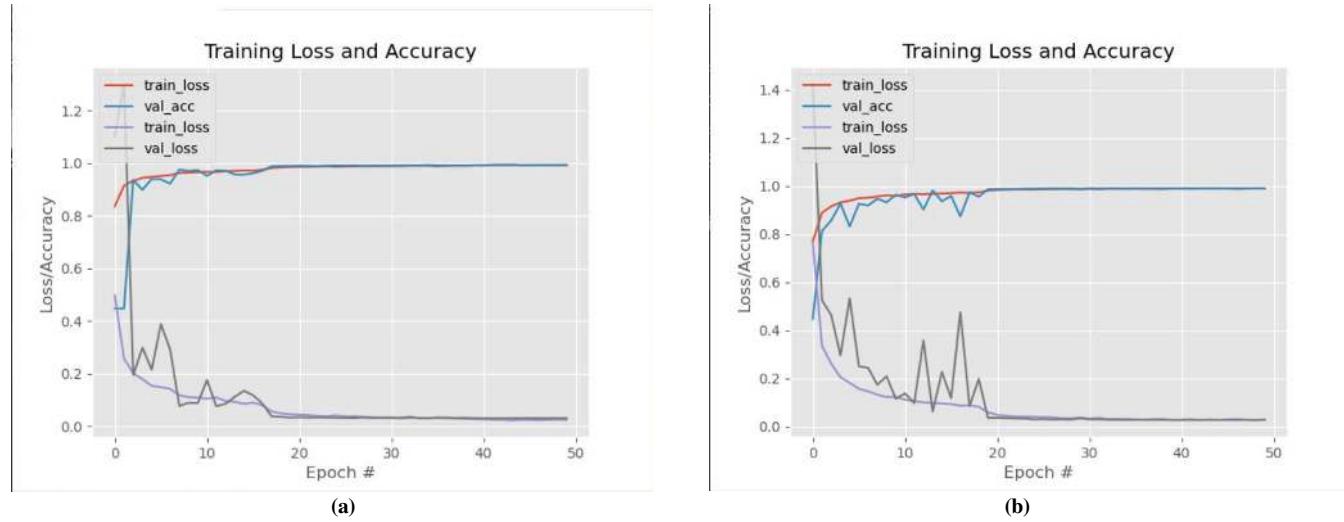


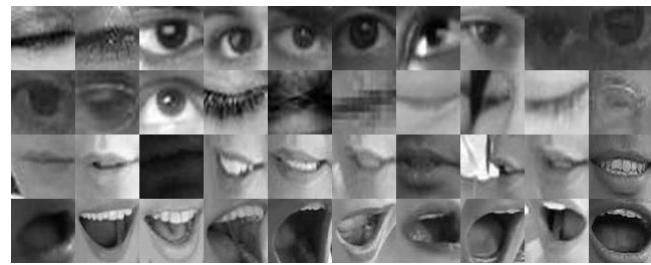
FIGURE 11. Test results of face location and facial key points detection module.
(a) Face location, and Eyes and mouth area detection; (b) Left eye area extraction; (c) Right eye area extraction; (d) Mouth area extraction

**FIGURE 13.** E-MSR Net model training curve. (a) MobileNetV2 training curve; (b) E-MSR Net training curve**TABLE 2.** Algorithm accuracy comparison

category	number	Accuracy of this algorithm	Accuracy of MobileNetV2
Closed-eye	324	98.9%	97.8%
Open-eye	392	99.8%	98.7%
Closed-mouth	2439	99.2%	99.1%
Open-mouth	2301	99.6%	98.3%

B. E-MSR NET MODEL TRAINING AND RESULT ANALYSIS

Using the self-built data set as the data set for E-MSR Net training. The data set contains a total of 21821 pictures of four types: open eyes, closed eyes, open mouth, and closed mouth, including 1500 pictures of open eyes, 1374 pictures of closed eyes, 9246 pictures of open mouth and 9701 pictures of closed mouth. Dividing the data set into training set, validation set, and test set at 7 : 1 : 2. The image of the data set section is shown in Figure 12. The experimental environment is the same as that of the previous experiment.

**FIGURE 12.** Eye and mouth state judgment module dataset

The training process is iterative 50 times, the first 20 times learning rate is set to 0.01, and the last 20 times learning rate is set to 0.001. The training process curve is shown in Figure 13. The algorithm in this paper is compared with the

MoblieNetV2 algorithm. It can be seen from the graph that the convergence speed of this algorithm is faster than MobileNetV2.

Using the test set to test the model trained by this algorithm and the model trained by MobileNetV2 respectively. The test results are shown in Table 2. The table shows that the accuracy of the E-MSR Net model trained is higher than that of the MobileNetV2 model.

C. FATIGUE DETECTION EXPERIMENT SIMULATION

The fatigue detection experiment uses self-made fatigue detection video as the test data. The test data includes fifty videos, each of which includes fatigue states such as eyes closed, yawning, and nodding. The detection effect is shown in Figure 14. The single evaluation index and the multi-feature fusion evaluation index are used to judge the fatigue state respectively, and the results of the comparison are shown in Table 3.

It can be seen from the figure that the algorithm in this paper can accurately and timely judge the fatigue state of the driver. Through the accuracy to measure the evaluation index of the algorithm, the formula is as follows:

$$\text{accuracy} = \frac{N_t}{N} \quad (9)$$

where N_t represents the number of fatigues driving states detected by the algorithm in this paper, and N represents the

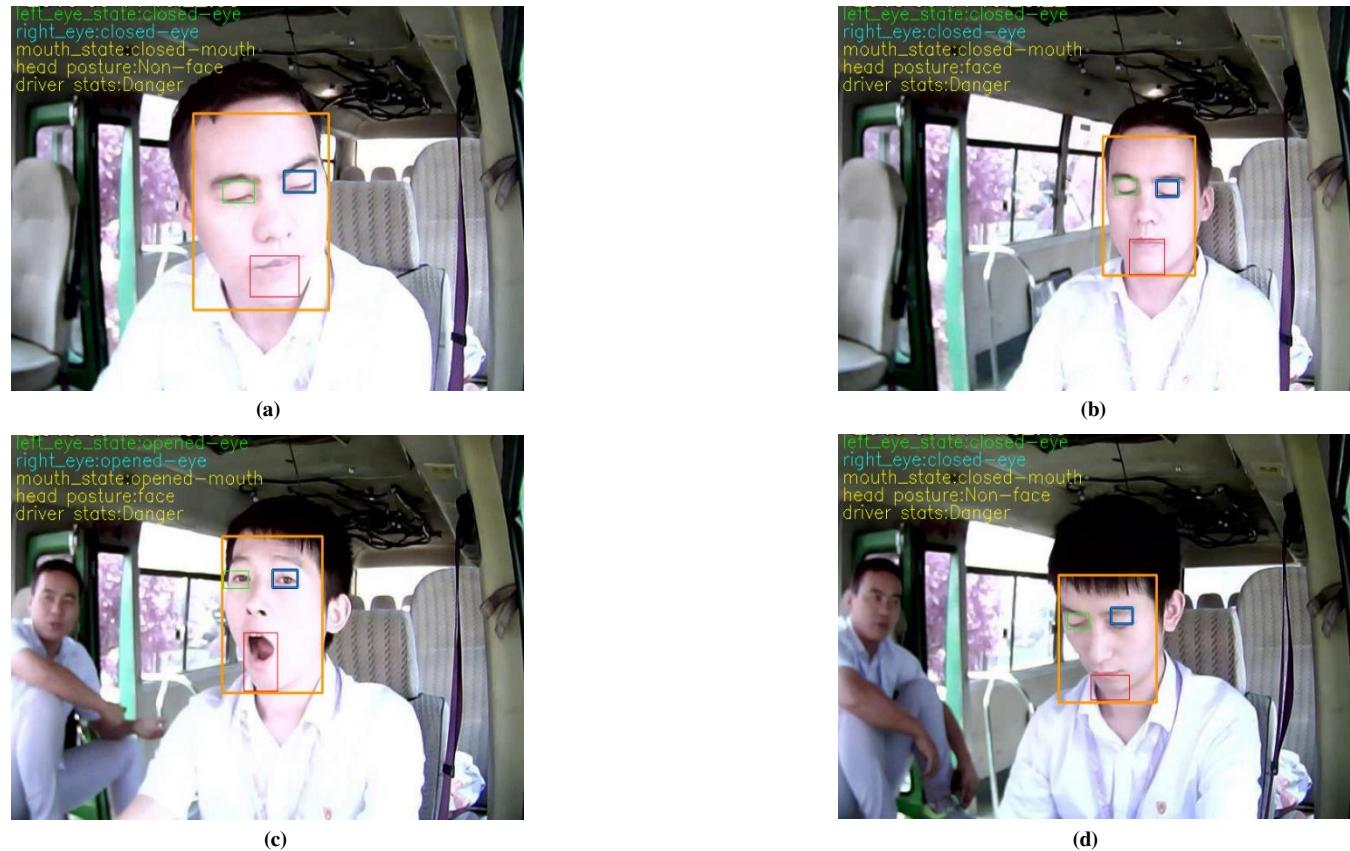


FIGURE 14. Test results of fatigue detection algorithm in this paper. (a) In a state of turning left and eyes closed; (b) In a state of closed eyes; (c) In a state of yawning; (d) In a state of bowed head and closed eyes.

TABLE 3. Comparison of single evaluation index and multi-feature fusion evaluation index

Fatigue characteristic parameters	Recognition results		Accuracy
	Fatigue	Normal	
f_e	36 times	14 times	90.0%
f_m	33 times	17 times	82.5%
f_h	30 times	20 times	75.0%
$f_e + f_m + f_h$	39 times	11 times	97.5%

number of fatigues driving states that really exist in the test video. Compare this algorithm with other algorithms, and the results are shown in Table 4.

TABLE 4. Comparison of different fatigue detection algorithms

Algorithm	Fatigue times	Accuracy
HOG+SVM [33]	35 times	87.5%
YOLOv3-tiny+SVM [34]	36 times	90.0%
Algorithm in this paper	39 times	97.5%

V. SUMMARY

The algorithm in this paper can deal with the fatigue detection of drivers in various complex environments. The improved MTCNN can accurately and quickly detect the face, eyes and mouth regions. E-MSR Net can accurately and quickly judge the state of mouth and eyes. The facial multi-feature fusion algorithm can accurately judge the fatigue state of drivers. The experimental results show that this algorithm can well detect different fatigue states.

REFERENCES

- [1] M. H. Alkinani, W. Z. Khan, and Q. Arshad, “Detecting human driver inattentive and aggressive driving behavior using deep learning: Recent

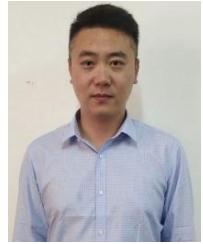
- advances, requirements and open challenges,” *IEEE Access*, vol. 8, pp. 105008–105030, 2020.
- [2] J. W. Jeong, H. K. Kim, Y. J. Lee, W. W. Jung, F. Harashima, and M. H. Lee, “The implementation of the autonomous guided vehicle driving system for durability test,” in *ITSC2000. 2000 IEEE Int. Transp. Syst. Proceedings (Cat. No.00TH8493)*, pp. 101–106, 2000.
- [3] R. Li, Y. V. Chen, and L. Zhang, “A method for fatigue detection based on driver’s steering wheel grip,” *Int. J. Ind. Ergonom.*, vol. 82, p. 103083, 2021.
- [4] M. Wang, N.-T. Jeong, K. Kim, S. B. Choi, S. M. Yang, S. You, J. Lee, and M. Suh, “Drowsy behavior detection based on driving information,” *Int. J. Automot.*, vol. 17, pp. 165–173, 2016.
- [5] M. Mao and L. Du, “Research on drive fatigue detection using wavelet transform,” in *2007 IEEE Int. Conf. Veh. Electron. Saf.*, pp. 1–4, 2007.
- [6] X.-Q. Huo, W.-L. Zheng, and B.-L. Lu, “Driving fatigue detection with fusion of eeg and forehead eog,” in *2016 Int. Joint Conf. Neural Netw. (IJCNN)*, pp. 897–904, 2016.
- [7] R. P. Balandong, R. F. Ahmad, M. N. Mohamad Saad, and A. S. Malik, “A review on eeg-based automatic sleepiness detection systems for driver,” *IEEE Access*, vol. 6, pp. 22908–22919, 2018.
- [8] L.-W. Ko, W.-K. Lai, W.-G. Liang, C.-H. Chuang, S.-W. Lu, Y.-C. Lu, T.-Y. Hsiung, H.-H. Wu, and C.-T. Lin, “Single channel wireless eeg device for real-time fatigue level detection,” in *2015 Int. Joint Conf. Neural Netw. (IJCNN)*, pp. 1–5, 2015.
- [9] M. S. Hossain, K. Huda, S. M. S. Rahman, and M. Ahmad, “Implementation of an eog based security system by analyzing eye movement patterns,” in *2015 Int. Conf. Adv. Elec. Eng. (ICAEE)*, pp. 149–152, 2015.
- [10] T. Kobayashi, S. Okada, M. Makikawa, N. Shiozawa, and M. Kosaka, “Development of wearable muscle fatigue detection system using capacitance coupling electrodes,” in *2017 39th Annual International Conference of the IEEE Eng. Med. Biol. Soc. (EMBC)*, pp. 833–836, 2017.
- [11] W. He, “Application heart rate variability to driver fatigue detection of dangerous chemicals vehicles,” in *2014 Fifth International Conference on Intelligent Systems Design and Engineering Applications*, pp. 218–221, 2014.
- [12] R. Bhardwaj, S. Parameswaran, and V. Balasubramanian, “Comparison of driver fatigue trend on simulator and on-road driving based on emg correlation,” in *2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS)*, pp. 94–97, 2018.
- [13] K. T. Chui, K. F. Tsang, H. R. Chi, B. W. K. Ling, and C. K. Wu, “An accurate ecg-based transportation safety drowsiness detection scheme,” *IEEE Trans. Ind. Informat.*, vol. 12, no. 4, pp. 1438–1452, 2016.
- [14] Y.-C. Tsai, P.-W. Lai, P.-W. Huang, T.-M. Lin, and B.-F. Wu, “Vision-based instant measurement system for driver fatigue monitoring,” *IEEE Access*, vol. 8, pp. 67342–67353, 2020.
- [15] B. K. Savas and Y. Bicerikli, “Real time driver fatigue detection system based on multi-task connn,” *IEEE Access*, vol. 8, pp. 12491–12498, 2020.
- [16] G. Soares, D. de Lima, and A. Miranda Neto, “A mobile application for driver’s drowsiness monitoring based on perclos estimation,” *IEEE Lat. Am. Trans.*, vol. 17, no. 02, pp. 193–202, 2019.
- [17] D. Liu, C. Zhang, Q. Zhang, and Q. Kong, “Design and implementation of multimodal fatigue detection system combining eye and yawn information,” in *2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP)*, pp. 65–69, 2020.
- [18] S. Ansari, F. Naghd, H. Du, and Y. N. Pahnwar, “Driver mental fatigue detection based on head posture using new modified relu-bilstm deep neural network,” *IEEE Trans. Intell. Transport. Syst.*, pp. 1–13, 2021.
- [19] E. Price, G. Moore, L. Galway, and M. Linden, “Towards mobile cognitive fatigue assessment as indicated by physical, social, environmental, and emotional factors,” *IEEE Access*, vol. 7, pp. 116465–116479, 2019.
- [20] W. Zhang and J. Su, “Driver yawning detection based on long short-term memory networks,” in *2017 IEEE Symp. Series Comput. Intell. (SSCI)*, pp. 1–5, 2017.
- [21] Q. Zhuang, Z. Kehua, J. Wang, and Q. Chen, “Driver fatigue detection method based on eye states with pupil and iris segmentation,” *IEEE Access*, vol. 8, pp. 173440–173449, 2020.
- [22] F. Zhang and F. Wang, “Exercise fatigue detection algorithm based on video image information extraction,” *IEEE Access*, vol. 8, pp. 199696–199709, 2020.
- [23] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, “Joint face detection and alignment using multitask cascaded convolutional networks,” *IEEE Signal Processing Lett.*, vol. 23, no. 10, pp. 1499–1503, 2016.
- [24] K. He, X. Zhang, S. Ren, and J. Sun, “Spatial pyramid pooling in deep convolutional networks for visual recognition,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 37, no. 9, pp. 1904–1916, 2015.
- [25] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *Proceedings of the 32nd International Conference on Machine Learning* (F. Bach and D. Blei, eds.), vol. 37 of *Proceedings of Machine Learning Research*, (Lille, France), pp. 448–456, PMLR, 07–09 Jul 2015.
- [26] A. Howard, A. Zhmoginov, L.-C. Chen, M. Sandler, and M. Zhu, “Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation,” in *CVPR*, 2018.
- [27] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *Proceedings of the IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, June 2018.
- [28] C.-H. Tsai, Y.-T. Chih, W. H. Wong, and C.-Y. Lee, “A hardware-efficient sigmoid function with adjustable precision for a neural network system,” *IEEE Trans. Circuits Syst. II: Express Briefs*, vol. 62, no. 11, pp. 1073–1077, 2015.
- [29] A. Acoglu, and E. Ercelebi , “Real time eye detection algorithm for perclos calculation,” in *2016 24th Signal Processing and Communication Application Conference (SIU)*, pp. 1641–1644, 2016.
- [30] 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.
- [31] H. Yang, L. Liu, W. Min, X. Yang, and X. Xiong, “Driver yawning detection based on subtle facial action recognition,” *IEEE Trans. Multimedia*, vol. 23, pp. 572–583, 2021.
- [32] S. Yang, P. Luo, C. C. Loy, and X. Tang, “Wider face: A face detection benchmark,” in *2016 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 5525–5533, 2016.
- [33] S. Dey, S. A. Chowdhury, S. Sultana, M. A. Hossain, M. Dey, and S. K. Das, “Real time driver fatigue detection based on facial behaviour along with machine learning approaches,” in *2019 IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON)*, pp. 28–30, 2020
- [34] K. Li, Y. Gong, and Z. Ren, “A fatigue driving detection algorithm based on facial multi-feature fusion,” *IEEE Access*, vol. 8, pp. 101244–101259, 2020.



JIA HUIJIE is currently a postgraduate student in School of Electrical Engineering and Automation, Qilu University of Technology (Shandong Academy of Sciences). His main research directions are computer vision and deep learning.



XIAO ZHONGJUN graduated from Shaanxi University of Science and Technology in 2011 with a doctorate. In 2011, he worked in the School of Electrical Engineering and Automation, Qilu University of Technology, and is currently the director of the Department of Automation in the School of Electrical Engineering. His main research direction are process industry automation, predictive control, and embedded system technology



JI PENG graduated from School of Instrumental Science and Engineering, Southeast University in 2017 and received a doctorate in engineering. In 2018, he worked in the School of Electrical Engineering and Automation, Qilu University of Technology, and is currently deputy director of the Department of Measurement and Control Technology and Instruments, School of Electrical Engineering and Automation. His main research directions are human-robot interaction technology, robot trajectory tracking control, visual servo control, active olfaction of mobile robots, target detection, recognition and tracking.