Comparison of the Deep Learning Methods Applied on Human Eye Detection

2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA) | 978-1-7281-9004-4/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICPECA51329.2021.9362685

Yunjie Xiang
School of Computer Science and Mathematics
Fujian University of Technology
Fuzhou, China
3284137586@qq.com

Haiyan Yang
School of Computer Science and Mathematics
Fujian University of Technology
Fuzhou, China
yanghy@fjut.edu.cn

Rong Hu
Fujian Provincial Key Laboratory of Automotive Electronics and Electric Drive
Fujian University of Technology
Fuzhou, China
hurong0910@qq.com

Chih-Yu Hsu
School of Computer Science and Mathematics
Fujian University of Technology
Fuzhou, China
Corresponding author: 61201903@fjut.edu.cn

***Abstract*—For the fatigue driving detection of a driver wearing a mask, the traditional fatigue driving detection method cannot effectively detect the face. The characteristics of the mouth area are disappeared due to the mask's occlusion. Therefore, the extraction of fatigue features in the eye area becomes very important. The accuracy of the eye area detection will directly affect the performance of the fatigue driving detection algorithm. At present, YOLOv3 and Faster-RCNN are both excellent models in the field of target detection. Therefore, this article uses the same data set and sets the same training parameters during training. Under a unified evaluation standard, the YOLOv3 model and the Faster-RCNN model are evaluated. Experimental results show that YOLOv3 has a better effect on human eye detection under the same conditions.**

Keywords—Fatigue driving, Mask, Eye detection, YOLOv3, Faster-RCNN

I. INTRODUCTION

Fatigue driving means that the driver has been driving for a long time or lacks sleep, and there is an imbalance between the physiological state and the psychological state [1-3]. Fatigue can cause distraction, narrow vision, and slow response to actions, which can lead to traffic accidents. According to statistics, about 20% of traffic accidents are caused by fatigue driving [4]. At present, the fatigue driving detection method has attracted more and more researchers' attention.

The main methods of fatigue driving detection can be divided into subjective detection methods and objective detection methods. The subjective methods are developed in the form of questionnaires. Famous questionnaire methods include Person Fatigue Scale, Cooper-Harper Assessment Questionnaire, Stanford Sleep Table, and Driving Record Table Wait.

The objective of fatigue driving detection methods include three types based on (1) driver physiological parameters (2) driver's facial features and (3) vehicle driving parameters. The fatigue detection method based on physiological characteristics mainly extracts the driver's EEG signal [5-8], ECG signal, and EMG signal [9] to judge fatigue. The fatigue detection method based on vehicle behavior information mainly detects the steering wheel steering angle, vehicle speed, lane offset [10] and other characteristic information to judge whether the driver is in fatigue state. The fatigue detection method based on visual features mainly uses the camera to shoot the driver in real-time and extracts the driver's eye-opening and closing state, blinking frequency [11,12], mouth opening degree [13,14] and head position [15] from the video to judge the fatigue. At present, the detection method based on the combination of visual features, vehicle running parameters, or physiological features has also been used [16,17]. However, only using visual features is relatively simple, only need a camera to complete fatigue detection, without other human body sensors or vehicle sensors and other equipment. Therefore, this paper uses the method based on visual features to detect fatigue. When the driver is in a state of fatigue, the facial features will change, and the driver's consciousness will be different from the conscious state. Therefore, the facial feature is an effective method to detect fatigue. Fatigue is judged by continuous acquisition and real-time analysis of the driver's facial feature data [18,19].

For the fatigue driving detection of drivers with masks, the features of the driver's mouth area cannot be detected due to the mask occlusion, and the face key points detection of drivers wearing masks is also affected.

Therefore, the driver's eye detection has become a very important issue. To extract the fatigue characteristics of the eye area, it is necessary to detect the eye area accurately and in real-time, and to distinguish the left and right eye categories. When the driver's head position changes, we can judge the deflection direction of the driver's head position according to the different symmetry of the left and right eyes and accurately distinguish the types of the left and right eyes. It is more convenient to extract the feature information of head position.

The current human eye detection methods can be divided

into indirect methods (that is, face detection first and then eye detection) and direct methods (eye detection directly in the image). In the indirect method, because it often depends on the face detection effect, it will lead to the failure of the eye detection task in the case of face detection failure. For example, under the occlusion condition of a mask, the effect of face detection is affected. In the direct method, the detection fails due to the small size of the collected human eye. For example, in gaze tracking and binocular vision systems, there are more cases where the human eye scale is too small. In addition, human eye detection must consider the real-time requirements of the system. For example, in fatigue driving detection, the detection efficiency is very important.

YOLOv3 [20,21] model and Faster-RCNN [22,23] model is both speed and precision target detection models. Therefore, this paper uses the YOLOv3 target detection model and Faster-RCNN target detection model to detect the driver's eye area. The two models are used to locate the eye area and classify the left and right eyes simultaneously. By using a unified evaluation index to compare the detection effect of the two models in the environment of mask occlusion.

II. Theory and Methods

This chapter mainly describes the theories and methods used in this research. In section A, the overall flow chart of the work of this article is described. The YOLOv3 target detection model is described in section B. The Faster-RCNN target detection algorithm is described in section C.

A. Flow chart

For the fatigue driving detection task of a driver wearing a mask, the accuracy of human eye detection directly affects the performance of the fatigue driving detection algorithm. Therefore, accurate, and fast human eye detection is the basic task of driving fatigue detection algorithm. At the same time, because of the particularity of the fatigue driving detection task, we need to choose a target detection algorithm that balances accuracy and real-time performance to achieve human eye detection. The workflow of this method is shown in Fig. 1. This method can overcome the interference of factors such as light in the driving environment and mask occlusion. It can accurately detect the position of the driver's eyes and the coordinates of the left and right eyes in real-time. Prepare for the subsequent extraction of eye fatigue features.

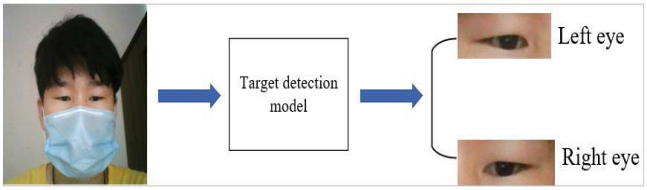

Fig. 1. The whole flow chart of this paper.

B. YOLOv3 target detection model

YOLOv3 is an improved version of the YOLO algorithm. YOLOv3 has made further improvements based on YOLOv1 and YOLOv2. The feature pyramid structure (FPN) based on multi-scale prediction is introduced into the network. Small objects will be detected in the shallow feature map, and large objects will be detected. It is detected in the deeper feature map.

The YOLOv3 network model is shown in Fig. 2, which is divided into three parts: feature extraction network, feature fusion layer, and output layer. YOLOv3 uses the DarkNet-53 network as the feature extraction network. Darknet-53 uses a fully convolutional network, which is composed of many 1×1 and 3×3 convolutional layers. The pooling layer is replaced by a convolutional layer with a step size of 2. The realization of downsampling not only reduces the amount of calculation when performing downsampling but also retains more information. At the same time, the Residual unit is added to avoid gradient dispersion when the network layer is too deep. In order to solve the problem that the previous YOLO version is not sensitive to small targets, the feature fusion layer of YOLOv3 uses 3 feature maps of different scales for target detection. The sizes are 13×13, 26×26, 52×52, respectively, for detection Three goals: large, medium, and small. The feature fusion layer selects three-scale feature maps produced by Darknet-53 as input and draws on the idea of FPN (feature pyramid networks). It combines feature maps of various scales through a series of convolutional layers and up-sampling layers. Finally, the output layer of YOLOv3 can get the position, size, and category information of the detection target in the image.

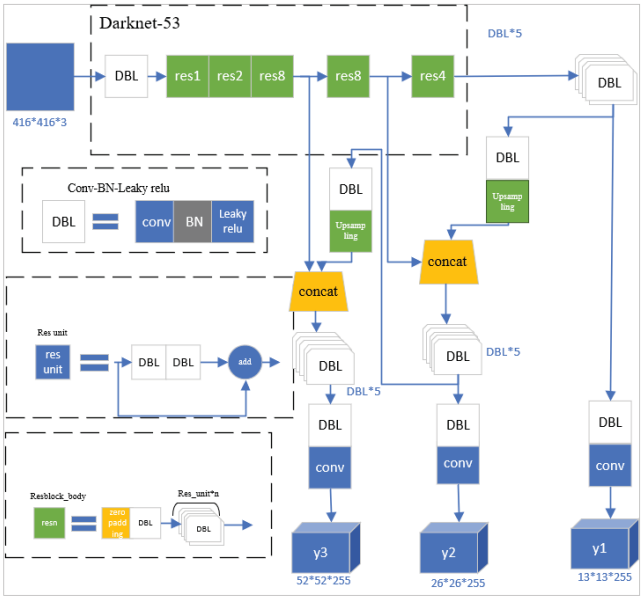

Fig. 2. Network structure diagram of YOLOv3.

C. Faster-RCNN target detection model

Faster-RCNN is also one of the best target detection models. It is built based on RCNN and Fast-RCNN. Faster-RCNN has made certain improvements in recognition accuracy and power consumption memory. It uses Region Proposal Networks (RPN) to replace the traditional sliding window and selective search (Selective Search) method to generate the detection frame. These improvements have greatly increased the detection speed, enabling real-time detection.

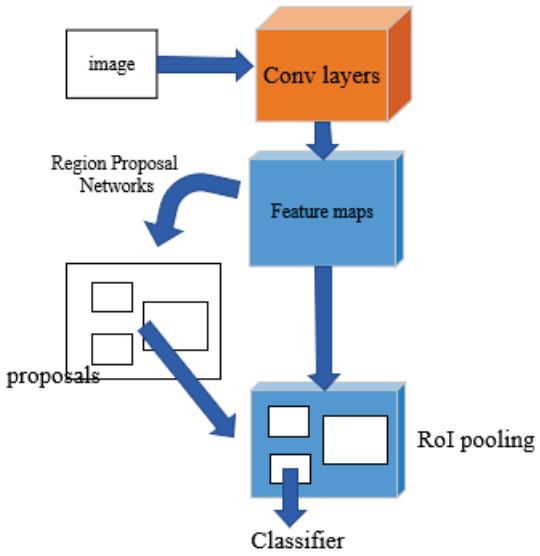

Fig. 3. Model structure diagram of Faster-RCNN.

As shown in Fig. 3, the Faster-RCNN model consists of four parts, including Conv layers, Region Proposal Networks (RPN), Proposal Layer, and Classification. Conv layers are the feature extraction part, which uses a series of convolution plus pooling to extract feature maps from the original image. The RPN part is a brand-new network structure proposed by Faster-RCNN. Its function is to obtain the approximate position of the target from the feature map through network training. The Proposal Layer part is to use the approximate position obtained by RPN and continue training to obtain a more accurate position. The ROI Pooling part uses the precise position obtained previously to extract the target to be used for classification from the feature map and pool it into fixed-length data.

III. Results and Discussion

In this chapter, the experimental data set and the experimental environment are introduced in Section A. In section B, the evaluation criteria used in this article are described. In section C, the experimental results of the YOLOv3 model and the Faster-RCNN model are described. The experimental discussion is carried out in section D.

A. Experimental data set and experimental environment

The data set used in this article consists of two parts, including the benchmark face data set WIDER FACE and the self-built face data set, with a total of 2783 images. The images in the data set are very different in scale, posture, occlusion, expression, dress, lighting, etc., which can effectively enhance the robustness of the trained model. The sample data in the data set is shown in Fig. 4.

The experiments in this research are all done on the windows operating system. The PaddlePaddle deep learning framework based on Baidu is used. This article uses wizard annotation assistant software for image annotation. In the face image, the category of the left eye is 0, and the category of the right eye is 1. In the following papers, we also use 0 for the left eye and 1 for the right eye. In the model training of this study, YOLOv3 and Faster-RCNN use the same data for training.

Fig. 4. Sample data in the dataset.

B. Evaluation criteria

To compare the performance of YOLOv3 and Faster-RCNN, this paper needs to use a unified evaluation standard to compare the two models. This paper uses the precision, the recall, the Precision-recall curve, and the mAP value to evaluate the performance of the model.

The precision indicates the proportion of the number of correctly detected eyes in all the test results. The calculation formula of the precision is as follows.

$$Precision = \frac{T_p}{F_p + T_p} \tag{1}$$

The recall indicates the proportion of the number of correctly detected eyes in all the tested samples. The formula for calculating the recall is shown below.

$$\text{Recall} = \frac{T_p}{T_p + F_n} \tag{2}$$

In formula (1) and formula (2), T_p (True positives) represents the number of positive samples correctly identified as positive samples. F_p (False positives) represents the number of negative samples that were incorrectly identified as positive samples. F_n (False negatives) represents the number of positive samples that are incorrectly identified as negative samples.

Among them, to obtain the mAP value, you first need to obtain the Precision-recall curve and calculate the average precision (AP). The Precision-recall curve represents the corresponding relationship between the precision rate and the recall rate. AP measures the quality of the trained model in each category. The AP calculation method adopts the 11-point method in VOC2007 (as shown in formula (4)). mAP is the average value of AP (as shown in formula (5)), and mAP reflects global performance.

$$P_{MaxPrecision}(R) = \max_{\tilde{R} \geq R} P(\tilde{R}) \tag{3}$$

$$AP = \frac{1}{11} \sum_{R \in (0, 0.1, \ldots, 1)} P_{MaxPrecision}(R) \tag{4}$$

$$mAP = \frac{1}{|Q_R|} \sum_{q \in Q_R} AP(q) \tag{5}$$

In formula (3), $P_{MaxPrecision}(R)$ is the maximum precision when the recall satisfies $\tilde{R} \geq R$, and Q_R is the number of categories. In this paper, the left eye and right eye categories are detected, so $Q_R = 2$.

C. Experimental results

This research first uses the YOLOv3 target detection model for human eye detection. The feature extraction network used by YOLOv3 is DarkNet-53. During model training, epochs=100, and learning rate=0.000125. After the training is completed, evaluate the trained model. And visualize the corresponding relationship between the accuracy rate and the recall rate of each category, and the corresponding relationship between the recall rate and the confidence threshold as shown in Fig. 5.

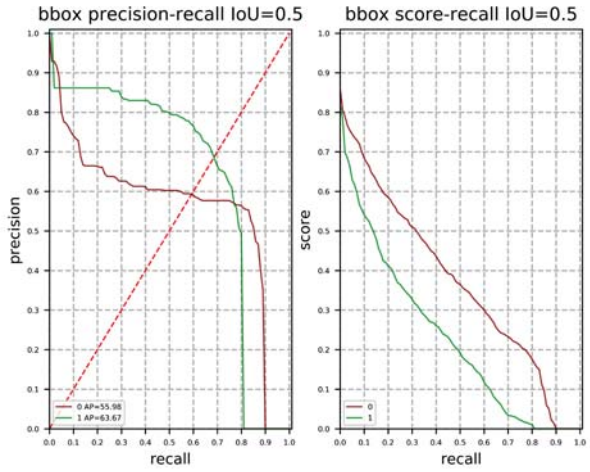

Fig. 5. The Precision-recall curve was obtained after the YOLOv3 model evaluation and the corresponding relationship between the recall and the confidence threshold.

Fig. 5 is the result of evaluating the performance of the YOLOv3 model. It shows the corresponding relationship between the accuracy rate and the recall rate of the eye category, and the corresponding relationship between the recall rate and the confidence threshold. The results show that when the confidence threshold is 0.5, it can be seen from the Precision-recall curve that the category is 0 and the AP value is 55.98. The AP value of category 1 is 63.67. It can be seen from the graph of the recall rate and score that the curve is showing a slow decline.

In this study, the Faster-RCNN target detection model is used for human eye detection. The feature extraction network used by Faster-RCNN is ResNet50. ResNet is the abbreviation of Residual Network. ResNet50 is a classic residual network, which is widely used in target classification and other fields and as a part of the classic neural network of the backbone of computer vision tasks. During model training, epochs=100, and learning rate=0.000125. After the training is completed, evaluate the trained model. And visualize the corresponding relationship between the precision rate and the recall rate of each category, and the corresponding relationship between the recall rate and the confidence threshold as shown in Fig. 6.

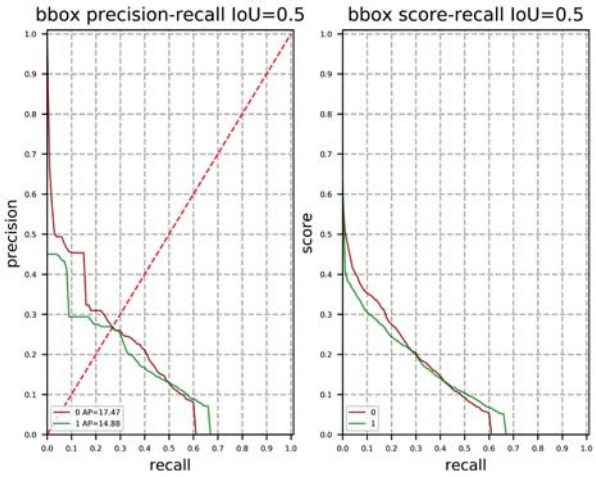

Fig. 6. The Precision-recall curve was obtained after the Faster-RCNN model evaluation and the corresponding relationship between the recall and the confidence threshold.

Fig. 6 is the result of evaluating the performance of the Faster-RCNN model. It shows the corresponding relationship between the accuracy rate and the recall rate of the eye category, and the corresponding relationship between the recall rate and the confidence threshold. The results show that when the confidence threshold is 0.5, it can be seen from the Precision-recall curve that the category is 0 and the AP value is 17.47. The AP value of category 1 is 14.88. It can be seen from the graphs of recall rate and score that the curve decline speed is relatively fast.

D. Discussions

The experimental results of YOLOv3 and Faster-RCNN were statistically analyzed. As shown in Table I, it can be seen from the table. Using the same data set, set the same training times, and learning rate. When detecting the driver's eye area, the AP value of the YOLOv3 model in category 0 was 55.98, and that in Category 1 was 63.67. The AP value of the Faster-RCNN model in category 0 was 17.47, and that in Category 1 was 14.88. Therefore, the performance of the YOLOv3 model is much better than that of the Faster-RCNN model in AP comparison. At the same time, the map value of the YOLOv3 model is 60.986302, and that of the Faster-RCNN model is 16.966874. Compared with Faster-RCNN, the performance of the YOLOv3 model is better than that of fast RCNN. Looking at Fig. 5 and Fig. 6, it is obvious that the precision-recall curve of the YOLOv3 model is above the Faster-RCNN model.

TABLE I. MODEL EVALUATION INDEX COMPARISON TABLE

index / model	0-AP	1-AP	mAP
YOLOv3	55.98	63.67	60.986302
Faster-RCNN	17.47	14.88	16.966874

IV. CONCLUSIONS

For the fatigue driving detection of a driver wearing a mask, the traditional fatigue driving detection method cannot effectively detect the face and extract the fatigue

characteristics of the mouth area due to the occlusion of the mask. Therefore, the feature extraction of the eye region becomes very important. The accuracy of human eye detection directly affects the performance of the fatigue driving detection algorithm. Therefore, accurate, and fast human eye detection is the basic task of driving fatigue detection algorithm. At the same time, because of the particularity of the fatigue driving detection task, we need to choose a target detection algorithm that balances accuracy and real-time performance to achieve human eye detection. This article uses the same data set and sets the same training parameters during training. Under a unified evaluation standard, the YOLOv3 model and the Faster-RCNN model are compared. Experimental results show that YOLOv3 has a better effect on human eye detection under the same conditions.

ACKNOWLEDGMENT

This research was completed with the support of the Fujian Provincial Department of Education Project (JA13219). The work is also supported by the Natural Science Foundation Fujian Province (No.2015J01652).

REFERENCES

[1] Guangnan Zhang, Kelvin K.W. Yau, Xun Zhang, and Yanyan Li, “Traffic accidents involving fatigue driving and their extent of casualties,” Accident Analysis & Prevention, Vol. 87, pp. 34-42, 2016.

[2] Sheng-Yang Shi, Wen-Zhong Tang, and Yan-Yang Wang, “A Review on Fatigue Driving Detection,” ITM Web Conf., Vol. 12, 01019, 2017.

[3] A. Amodio, M. Ermidoro, D. Maggi, S. Formentin, and S. M. Savaresi, "Automatic Detection of Driver Impairment Based on Pupillary Light Reflex," in IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 8, pp. 3038-3048, Aug. 2019, doi: 10.1109/TITS.2018.2871262.

[4] Ralston Fernandes, Julie Hatfield, and R.F. Soames Job, “A systematic investigation of the differential predictors for speeding, drink-driving, driving while fatigued, and not wearing a seat belt, among young drivers,” Transportation Research Part F: Traffic Psychology and Behaviour, Vol. 13, no. 3, pp. 179-196, 2010, doi.org/10.1016/j.trf.2010.04.007.

[5] Hong Wang, Chi Zhang, Tianwei Shi, Fuwang Wang, and Shujun Ma, “Real-Time EEG-Based Detection of Fatigue Driving Danger for Accident Prediction,” International Journal of Neural Systems, Vol. 25, no. 02, 1550002, 2015, doi.org/10.1142/S0129065715500021.

[6] Z. Gao et al., "Relative Wavelet Entropy Complex Network for Improving EEG-Based Fatigue Driving Classification," in IEEE Transactions on Instrumentation and Measurement, vol. 68, no. 7, pp. 2491-2497, July 2019, doi: 10.1109/TIM.2018.2865842.

[7] Haowen Luo, Taorong Qiu, Chao Liu, and Peifan Huang, “Research on fatigue driving detection using forehead EEG based on adaptive multi-scale entropy,” Biomedical Signal Processing and Control, Vol. 51, pp. 50-58, 2019, doi.org/10.1016/j.bspc.2019.02.005.

[8] Difei Jing, Dong Liu, Shuwei Zhang, and Zhongyin Guo, “Fatigue driving detection method based on EEG analysis in low-voltage and hypoxia plateau environment,” International Journal of Transportation Science and Technology, vol. 9, no. 4, pp. 366-376, 2020, doi.org/10.1016/j.ijtst.2020.03.008.

[9] L. Boon-Leng, L. Dae-Seok and L. Boon-Giin, "Mobile-based wearable-type of driver fatigue detection by GSR and EMG," TENCON 2015 - 2015 IEEE Region 10 Conference, Macao, 2015, pp. 1-4, doi: 10.1109/TENCON.2015.7372932.

[10] F. Friedrichs and B. Yang, "Drowsiness monitoring by steering and lane data based features under real driving conditions," 2010 18th European Signal Processing Conference, Aalborg, 2010, pp. 209-213.

[11] K. Mukherjee, R. Karmakar and S. Das, "Effective Estimation of Driver Drowsiness Based on Eye Status Detection and Analysis," 2014 International Conference on Devices, Circuits and Communications (ICDCCom), Ranchi, 2014, pp. 1-4, doi: 10.1109/ICDCCom.2014.7024717.

[12] Zhongmin Liu, Yuxi Peng, and Wenjin Hu, “Driver fatigue detection based on deeply-learned facial expression representation,” Journal of Visual Communication and Image Representation, Vol. 71, 2020, no. 102723, doi.org/10.1016/j.jvcir.2019.102723.

[13] Chang Zheng, Ban Xiaojuan, and Wang Yu, “Fatigue driving detection based on Haar feature and extreme learning machine,” The Journal of China Universities of Posts and Telecommunications, Vol. 23, no. 4, pp. 91-100, 2016, doi.org/10.1016/S1005-8885(16)60050-X.

[14] K. Li, Y. Gong and Z. Ren, "A Fatigue Driving Detection Algorithm Based on Facial Multi-Feature Fusion," in IEEE Access, vol. 8, pp. 101244-101259, 2020, doi: 10.1109/ACCESS.2020.2998363.

[15] W. Deng, Z. Zhan, Y. Yu and W. Wang, "Fatigue Driving Detection Based on Multi Feature Fusion," 2019 IEEE 4th International Conference on Image, Vision and Computing (ICIVC), Xiamen, China, 2019, pp. 407-411, doi: 10.1109/ICIVC47709.2019.8980929.

[16] Xu Li, Lin Hong, Jian-chun Wang, Xiang Liu, “Fatigue driving detection model based on multi-feature fusion and semi-supervised active learning,” IET Intelligent Transport Systems, Vol. 13, no. 9, pp. 1401-1409, 2019, doi: 10.1049/iet-its.2018.5590.

[17] J. Ma, J. Zhang, Z. Gong and Y. Du, "Study on Fatigue Driving Detection Model Based on Steering Operation Features and Eye Movement Features," 2018 IEEE 4th International Conference on Control Science and Systems Engineering (ICCSSE), Wuhan, China, 2018, pp. 472-475, doi: 10.1109/CCSSE.2018.8724836.

[18] Y. Qiao, Kai Zeng, Lina Xu and Xiaoyu Yin, "A smartphone-based driver fatigue detection using fusion of multiple real-time facial features," 2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, 2016, pp. 230-235, doi: 10.1109/CCNC.2016.7444761.

[19] F. Zhang, J. Su, L. Geng and Z. Xiao, "Driver Fatigue Detection Based on Eye State Recognition," 2017 International Conference on Machine Vision and Information Technology (CMVIT), Singapore, 2017, pp. 105-110, doi: 10.1109/CMVIT.2017.25.

[20] Zhang Yi, Shen Yongliang, and Zhang Jun, “An improved tiny-yolov3 pedestrian detection algorithm,” Optik, Vol. 183, pp. 17-23, 2019, doi.org/10.1016/j.ijleo.2019.02.038.

[21] Yunong Tian, Guodong Yang, Zhe Wang, Hao Wang, En Li, and Zize Liang, “Apple detection during different growth stages in orchards using the improved YOLO-V3 model,” Computers and Electronics in Agriculture, Vol. 157, pp. 417-426, 2019, doi.org/10.1016/j.compag.2019.01.012.

[22] Xudong Sun, Pengcheng Wu, and Steven C.H. Hoi, “Face detection using deep learning: An improved faster RCNN approach,” Neurocomputing, Vol. 299, pp. 42-50, 2018, doi.org/10.1016/j.neucom.2018.03.030.

[23] X. Zhao, W. Li, Y. Zhang, T. A. Gulliver, S. Chang and Z. Feng, "A Faster RCNN-Based Pedestrian Detection System," 2016 IEEE 84th Vehicular Technology Conference (VTC-Fall), Montreal, QC, 2016, pp. 1-5, doi: 10.1109/VTCFall.2016.7880852.