Early warning system for drivers' phone usage with deep learning network

J. H. Jixu Hou[1], Xiaofeng Xie[2], Qian Cai[1], Zhengjie Deng[1], Houqun Yang[2], Hongnian Huang[3], Xun Wang[4], Lei Feng[4] and Yizhen Wang[1*]

*Correspondence:
wangxuesu1980@163.com
[1] College of Physics &
Electronic Engineering,
Hainan Normal University,
Haikou, China
Full list of author information
is available at the end of the
article

Abstract

Dangerous driving, e.g., using mobile phone while driving, can result in serious traffic problem and threaten to safety. To efficiently alleviate such problem, in this paper, we design an intelligent monitoring system to detect the dangerous behavior while driving. The monitoring system is combined by a designed target detection algorithm, camera, terminal server and voice reminder. An efficiently deep learning model, namely Mobilenet combined with single shot multi-box detector (Mobilenet-SSD), was applied to identify the behavior of driver. To evaluate the performance of proposed system, a dangerous driving dataset,consisting of 6796 images, was constructed. The experimental results show that the proposed system can achieve the accuracy of 99%, and could be used for real-time monitoring of the drivers' status.

Keywords: Deep learning, Driving safety, Object detection, Computer vision, Machine learning

1 Introduction

With increasing of private car, the traffic accident rate is also rising. The inattention caused by mobile phones in driving is one of the main reason for traffic accident [1, 2]. By detecting the drivers' behavior of playing phone, it is of great significance to prevent traffic accidents from the perspective of the driver. Therefore, how to efficiently detect the drivers' behavior in driving has attracted increasing attention in recent years.

At present, the detection methods of drivers' mobile phone playing are mainly divided into on-site law enforcement and automatic detection by law enforcement officers [1–4]. However, the two monitoring methods are not efficient enough and can not be widely used in all private cars. Thus, in this work we develop the auto-monitor system in car to detect the drivers' behavior in driving.

In recent year, some related researches have been proposed to detect the behavior of driver based on machine learning method. In previous studies [3, 4], an activity parsing algorithm was employed to identify whether the driver was using a mobile phone. It used the And-Or Graph (AoG) to represent the hierarchically compositions of the phoning activity to efficiently improve the detection performance.A driver phone calls detection based on voice feature recognition was proposed [4]. It could recognize the drivers'

Hou *et al. J Wireless Com Network*   (2022) 2022:42

Page 2 of 11

voice from the collected audio data and determine whether the driver was participating in the current phone call or not. In Yasar's workb5, a neural network application was used to detect mobile phone usage with a outside camera. Berri and Osorio [6] developed a 3D vision system, using a frontal Kinect v2 sensor, to monitor the driver and monitor the use of mobile phones by drivers. Some studies [7] show that the drivers' face region could be localized by using the deformable part model (DPM), and a local aggregation based image classification technique could be utilized to classify a region of interest (ROI) around the drivers face for detecting the cell phone usage. Moreover,the Supervised Descent Method (SDM) [8] based facial landmark tracking algorithm was demonstrated to be able to track the position of face landmarks for detection,thus to determine if a driver is holding a cell phone in dirving. Despite the previous studies had made some progress, the influence due to the interference from the environment, e.g., different light and irrelevant background, was ignored, and the detection area in most previous studies was limited to face area only. All these limits the application of previous methods [1–8]. Recently, it has shown that the interference from the environment has a great impact on safe driving detection [8].

In this work, we construct a lightweight deep network model to detect the behavior of playing mobile phone in the complex environment. The proposed monitor system is divided into two parts: the vehicle mobile terminal and the PC terminal. The vehicle mobile terminal quickly detects the drivers' pictures collected by the in-car camera. The PC terminal can automatic communicate with the government department in the case of the drivers' phone usage. The major contributions of our whork are:

1. A novel deep network is proposed for detecting drivers' mobile phone behavior. Compared with traditional machine learning method in previous works [1–8], we proposed a mobilenNet combined with the single shot multi-box detector (mobileNet-SSD) to achieve object detection. It is a lightweight network that can fulfill a request of practical application.
2. The performance of proposed monitor system can achieve high performance for behavior recognition. The experimental results show that the proposed can achieve 99% accuracy in detection.
3. The detection results can sent to government department and help the traffic police to check the illegal behaviors of playing mobile phones.

This paper is organized as follows: In Sect. 1, the research backgroud is presented. In Sects. 2 and 3, the construction of the early warning system for drivers' playing mobile phone, as well as the details and functional of each module of monitor system are presented. Sect. 4 gives the experimental results and correspondig discussion. The conclusion are provided in Sect. 5.

2 Methods

2.1 The framework of early warning system

The framework of early warning system for detecting the drivers' dangerous behavior is shown in Fig. 1. It is composed of the mobile terminal and PC part. In the vehicle mobile terminal, the target detection technology in the field of computer vision is used
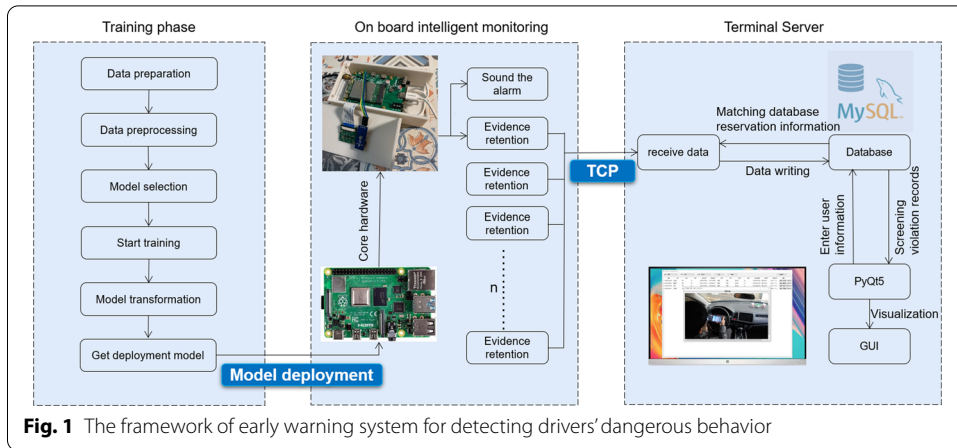
Fig. 1 The framework of early warning system for detecting drivers' dangerous behavior
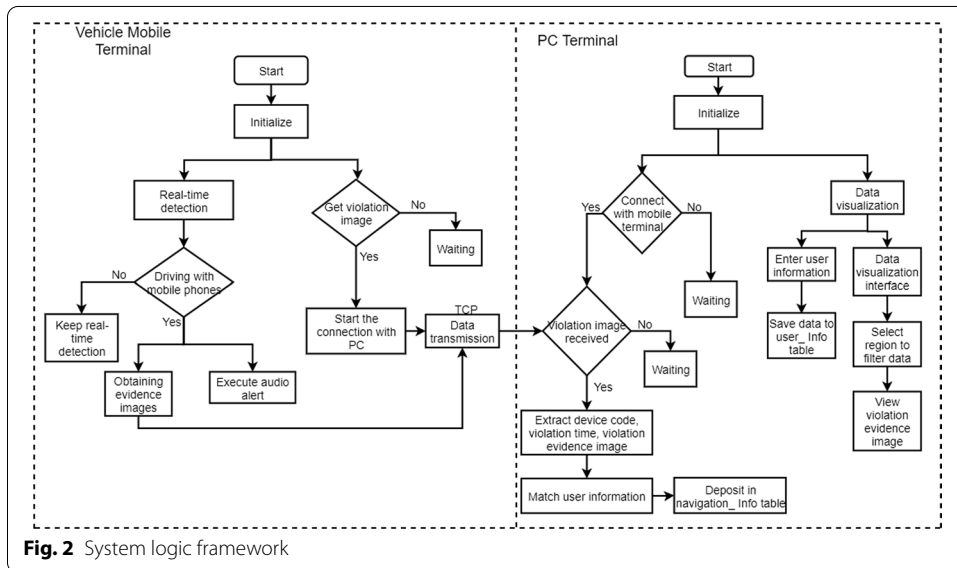

Fig. 2 System logic framework

to train the sample image model, and deploy the trained model on raspberry pi, the raspberry pi development board is configured to combine with the camera, 4G LTE module, Bluetooth speaker and other hardware, the TCP connection with the server is realized through 4f LTE module to transmit data. In the PC part, the technology of information input, recognition data wireless receiving, data statistics recording and data visualization are used to achieve the real-time status monitoring of the driver.

2.2 The implementation process of system

The implementation of the system is divided into two parts, vehicle mobile terminal and PC terminal. They communicate by TCP technology. The on-board mobile terminal monitors can detect dangerous driving behavior by executing the real-time monitoring module. If the system detects the phone's behavior, it will send an alarm and store the violation evidence image, which will be uploaded to the PC terminal database for storage. After receiving the violation evidence image, the PC terminal will compare the information and stores it in the corresponding location. As shown in Fig. 2, the detailed

Hou *et al. J Wireless Com Network*   (2022) 2022:42

Page 4 of 11

implementation process is subdivided into 4 modules, e.g., audio alarm module, wireless feedback data receiving module, database statistical recording module, and data visualization module.

2.2.1 Audio alarm module

Realization function: The framework of audio alarm module is shown in Fig. 3. Many people will inadvertently pick up the phone, regardless of driving safety. This product is expected to give an alarm after the detection of mobile phone target to remind the driver to regulate driving [9]. Therefore, an audio alarm module is added. When the real-time monitoring module detects that the driver is playing mobile phone, it will give an alarm to remind the driver to regulate driving behavior.

   Scheme design: The active buzzer module triggered by high level is used as the sound device. When the mobile phone is detected, the GPIO interface of the main board is triggered to output high level for sound alarm.

### 2.2.2 Wireless receiving module of feedback data

Realization function: PC terminal receive the image evidence and relevant time data from each on-board detection equipment.

   Scheme design: After initializing port and IP address, it use to capture socket after successful connection. Then it get the content of message through recv and close the socket after communication finish. The framework of wireless receiving module is shown in Fig. 4.

2.2.3 Database statistics record module

Realization function: In the process of driving, the traffic cameras on the road have fixed positions, which can not capture the evidence of violation in real time. We hope that through this system, we can record the illegal operation behavior and transmit it to the traffic department. Therefore, we add the database statistical record module to record the illegal operation behavior when we find playing mobile phones. If the correction fails, the information of vehicle owner and vehicle can be recorded and the evidence of violation can be stored.

Fig. 3 Work flow of audio alarm module

Fig. 4 The framework of wireless receiving module

Fig. 5 The framework of database module
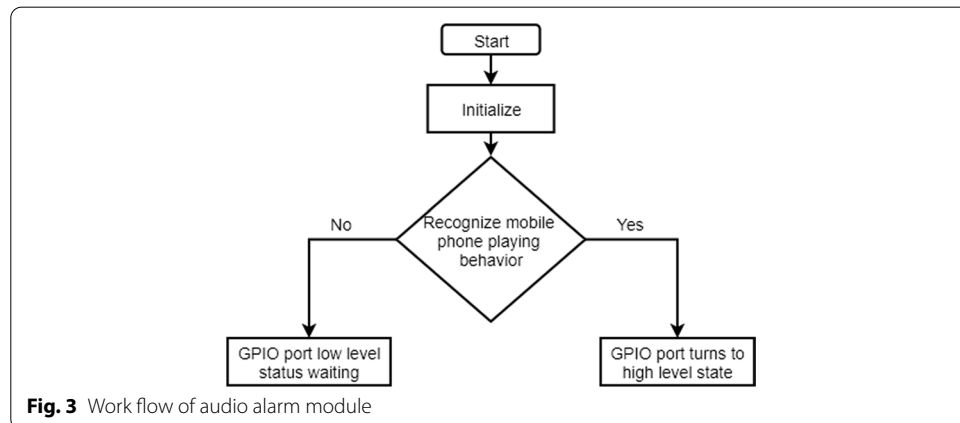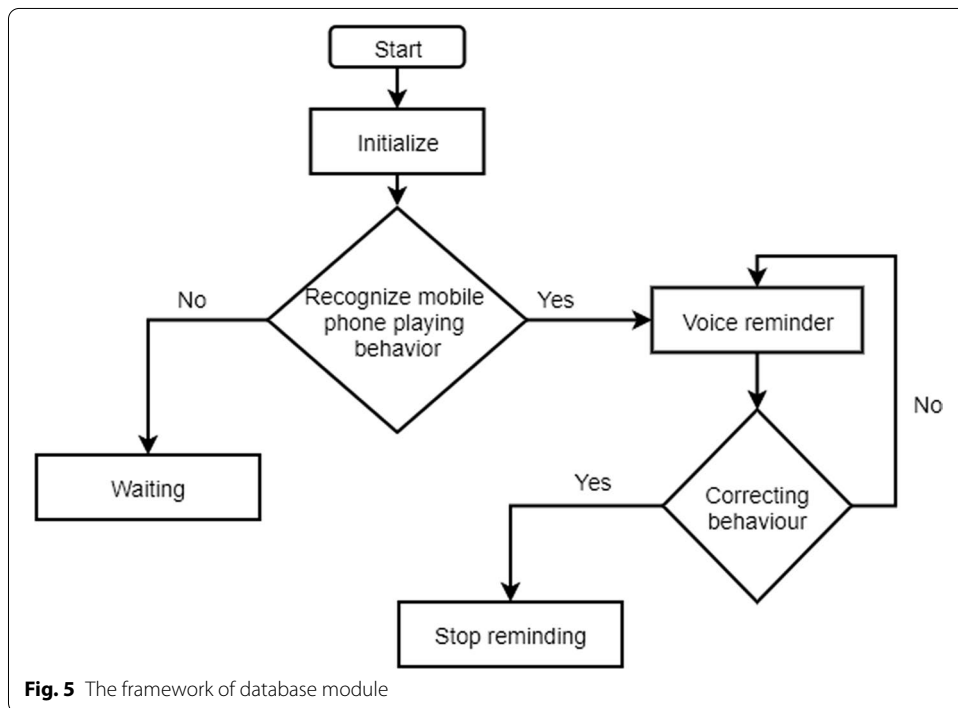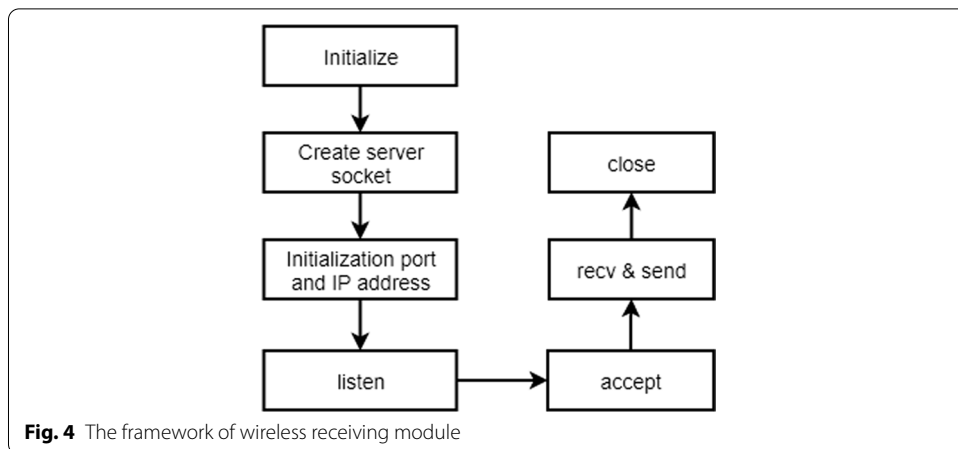
Scheme design: It use MySQL database to create a data table for data storage, including owner and vehicle information. It can capture violation time, image evidence and other data. The framework of database module is shown in Fig. 5.

2.2.4 Data visualization module

Realization function: In the actual work process, it is not convenient to directly operate the database to view information. Therefore, it is necessary to display the data in the database by a visual operation interface, and display the evidence of the drivers' violation.

Scheme design: By using pyqt5, we design main interface to display urban information, and to view the illegal image.

Hou *et al. J Wireless Com Network*   (2022) 2022:42

Page 6 of 11

3 Deep learning method for detection

3.1 Model training data

3.1.1 *Data acquisition*

In order to ensure the accuracy of mobile phone recognition and make the model better applied to the actual work scene, this paper collects image data in different automotive interior environment, as illustrated in Fig. 6. And we also expand the image data set by flipping, mirroring, and clipping. As results, we obtains 6796 image data for training model. All images are marked with labelimg tool.

3.1.2 *Data preprocessing*

Due to the bumpy car body and lack of light during the driving process, the captured image may also have problems such as shaking and blurring, which easily makes the image blurred and lacks the characteristics of the corresponding target, and reduces the quality of image data. Therefore, before training, the image data preprocessing mainly includes Gaussian blur, edge enhancement, etc. In the process of training, it can learn the characteristics of mobile phone in more complex scenes.

3.2 Detection algorithm

In this paper, we used the mobilenNet combined with the single shot multi-box detector (mobilenNet-SSD) [10] for behavior detection. Compared with YOLO [11] and Faster-RCNN [12], the mobilenNet-SSD algorithm can utilize different size boxes to regression at all pixels of the whole picture. For Faster RCNN [12], it needs to get the bounding box through CNN before classification and regression. However, YOLO [11] and SSD [10] can complete the detection in one stage. Compared with YOLO [11], SSD [10] uses CNN to detect directly, rather than using the full connection layer as YOLO [11]. And SSD [10] extracts feature maps of different scales for detection. In this paper, the mobile
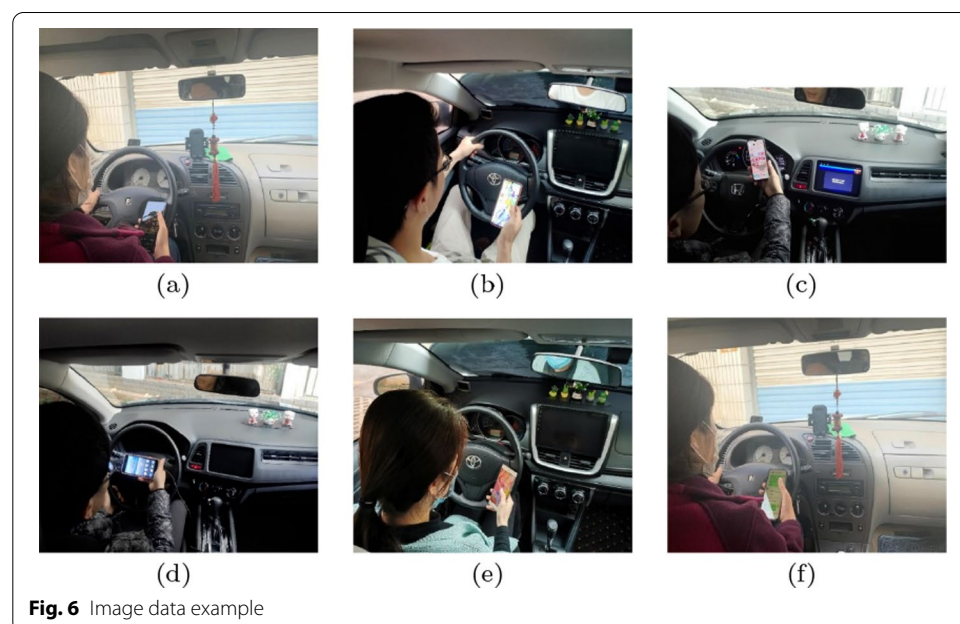

Fig. 6 Image data example

phone target size is relatively fixed and the features are not complex. Therefore, we use the large-scale feature map of SSD [10] algorithm to detect small objects, small feature map to detect the characteristics of large objects, and delete the detection of two large feature maps, so as to further improve the balance between speed and accuracy. On this foundation, we replace the basic network with mobilenet_v3(small) [13, 15]. In the mobilenet_v3 [13, 15] structure, the original author use depthwise separable convolution to reduce the number of parameters [11]. Therefore, this design is more suitable for small mobile devices with limited computing power.

3.2.1 Algorithm adjustment

The detection module uses ssdlite_mobilenet_v3_small network implementation. Instead of ssdlite [14], convolution in SSD is replaced by deep separable convolution. The mobilenet_V3(small) [15] is used as the basic network and the basic network is deleted from the 2-layer feature map. The lightweight network collocation can greatly reduce the computation, improve the detection effect and speed of the model, and complete the high-speed identification, and warning work of mobile devices.

3.2.2 Model training

In order to ensure the accuracy of the trained model, transfer learning is used in this training, which transplants the parameters from the pre-training model. Since pre-training model has been trained and performed well, a more accurate model can be obtained in the case of a small data set, and the over fitting phenomenon will not be caused by the small data scale.

3.2.3 Model performance

The collected image data is labeled with mobile phone position and trained. The trained model is converted into prediction model. The function is realized by using the lightweight reasoning engine of paddlelite deployed to the vehicle mobile terminal. Figure 7 illustrates the detection result image.

In the optimization and improvement of the algorithm, we improve the connection between mobilenet_v3_small and ssdlite .In the original algorithm, the default boxes of ssdlite_ mobilenet_v3_small are generated from the feature maps output by the six convolution layers. In this paper, we abandon the two convolution layers for detecting small
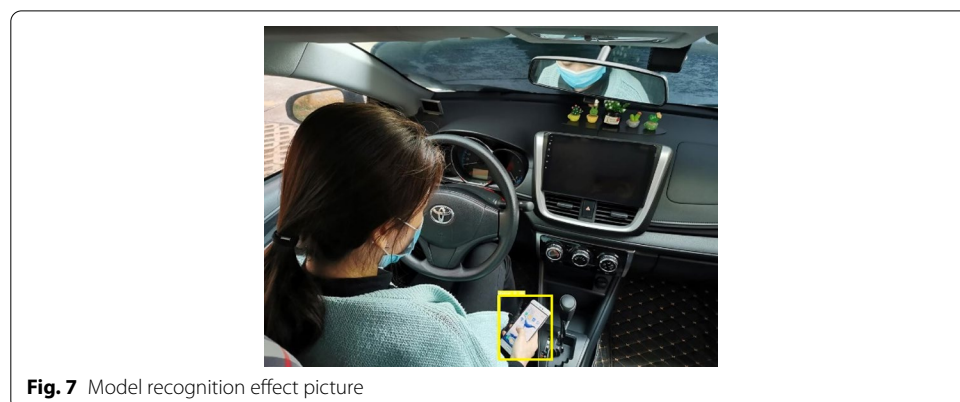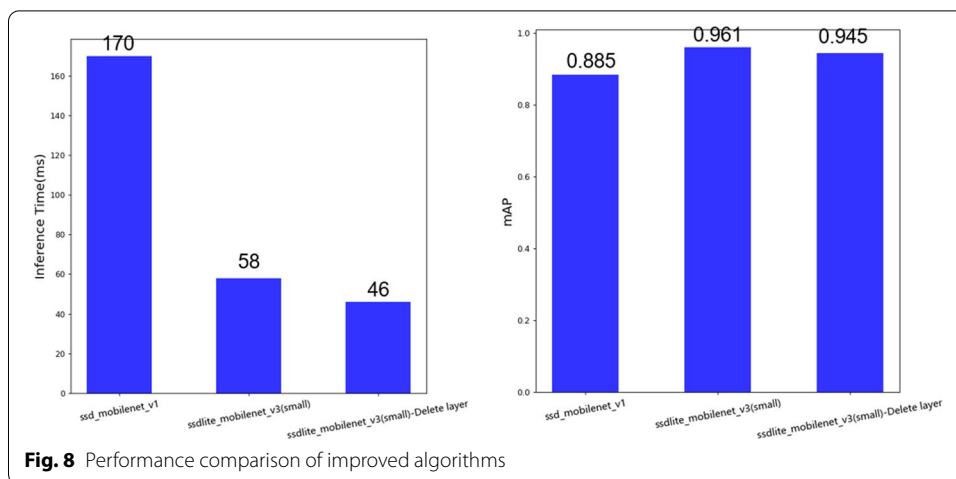

Fig. 7 Model recognition effect picture

Hou *et al. J Wireless Com Network*   (2022) 2022:42

Page 8 of 11

Fig. 8 Performance comparison of improved algorithms

Table 1 Pattern recognition data

Accuracy	mAP (IoU = 0.5)	mAP (IoU = 0.75)	Inference time (ms)
0.99	0.997	0.945	46

targets and only use the feature map output by the four convolution layers for calculation. For simple targets such as mobile phones, it can not only maintain the accuracy, but also improve the detection efficiency. We compared three models: ssd_mobilenet_v1, ssdlite_mobilenet_v3(small) and the performance of ssdlite_mobilenet_v3(small) after layer deletion. Figure 8 shows the accuracy and inference time of three methods. As shown in Fig. 8, we can see the time consumption of SSD_mobilenet_v1 model is about three times that of ssdlite_mobilenet_v3(small) model. After the deletion operation, the speed is increased by about 20.7%, and the accuracy does not change much. By comparing the accuracy and speed of the model, mobilenet_ssd model has been improved, which effectively improves the running speed of the model, and can detect and judge more quickly and accurately.

4 Results and discussion

Equipment response efficiency is a key factor when designing them. Specifically, if we want to remind drivers to correct driving behavior in time, we need to improve the recognition efficiency. In this regard, we need to correctly analyze ssdlite_mobilenet_v3(small) algorithm, and our recognition target is relatively simple. We can reduce unnecessary calculation by simplifying the model.

Table 1 shows the proposed method can achieve 99% accuracy in 100 images. And IoU represents the ratio of intersection and union of prediction frame and real frame. While IoU is set to 0.5, the accuracy achieve 99.7%. However while IoU is set to 0.75, the accuracy is 94.5%; The inference time is 46 ms.

We also compared the proposed method with the Yolov3 [16] and Faster-RCNN [12] network. As shown in Table 2, the accuracy of the three methods is high, but the running times of three method are different. The proposed network achieve 46 ms, while Yolov3 [16] network cost 4799.8 s, and the Faster-RCNN is demonstrated to be

Hou *et al. J Wireless Com Network*   (2022) 2022:42

Page 9 of 11

Table 2 Comparison of inference time and accuracy of different models

Model	Inference time (ms)	Accuracy (%)
ssd_mobilenet_v3(small)	46	99
Yolov3	4799.8	99
Faster-RCNN	–	95

not applicable to raspberry pi used in this system due to its high computational cost [12]. Moreover, the ssd_mobilenet_v3 and Yolov3 models were demosntrated to have the better accuracies used in this system than that of the Faster-RCNN, as shown in Table 2. Note the accuracy of corresponding Faster-RCNN shown in Table 2 was achieved in the AI Studio server [17].

4.1 Physical appearance design

In this paper, we designed the physical appearance of the system ourselves and printed it with a 3D printer (ANYCUBIC Chiron 3D). As shown in Figs. 9 and 10, the internal circuit and physical appearance are designed. The whole product is small and convenient. It occupies a small space and easy to install.

Fig. 9 Internal circuit display

Fig. 10 Physical appearance display
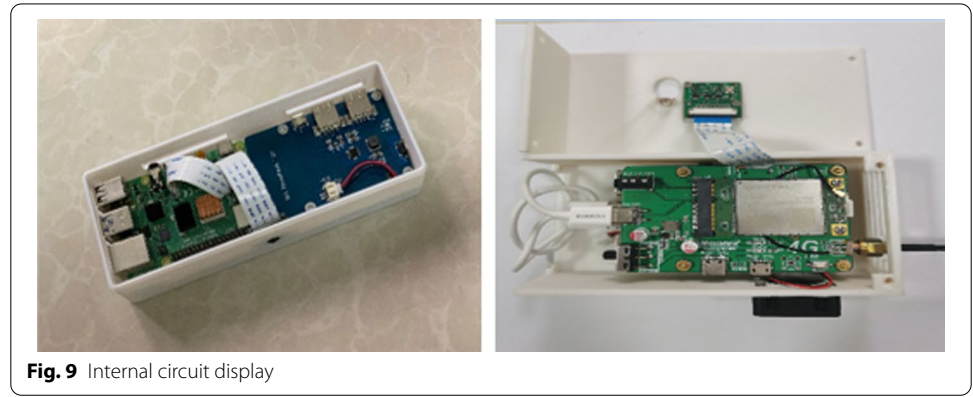
5 Conclusions

In this paper, we design a drivers' phone usage detection system. It is composed of the mobile terminal and PC part. It used mobilenNet combined with the single shot multi-box detector to achieve object detection. Compared with other deep network, the proposed model can achieve high classification performance with less computational cost. It is a lightweight network that can fulfill a request of practical application. The proposed system can also applied in other detection fields, e.g., fatigue driving or driving without seat belt.

Declarations

Competing interests
The authors declare that they have no competing interests.

Author details
[1]College of Physics & Electronic Engineering, Hainan Normal University, Haikou, China. [2]College of Mechanical & Electrical Engineering, Hainan University, Haikou, China. [3]Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM, USA. [4]Hainan Blue Dot Computer Network Engineering Co., Ltd, Haikou, China.

References
1. P. Hancock, M. lesch, L. Simmons, The distraction effects of phone use during a crucial driving maneuver. Accid. Anal. Prev. **35**(4), 501–514 (2003)
2. W. Li, Pay attention to fleet safety management to reduce traffic accidents. Commer. Veh. **000**(021), 105–107 (2012)
3. D. Wang, M. Pei, L. Zhu, Detecting driver use of mobile phone based on in-car camera. In: Tenth International Conference on Computational Intelligence and Security (2014), pp. 148–151. https://doi.org/10.1109/CIS.2014.12
4. T. Song, X. Cheng, H. Li, J. Yu, S. Wang, R. Bie, Detecting driver phone calls in a moving vehicle based on voice features. In: IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications, (2016), pp. 1–9. https://doi.org/10.1109/INFOCOM.2016.7524437
5. H. Yasar, Detection of driver's mobile phone usage. In: IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM) (2017), pp. 1–4. https://doi.org/10.1109/HNICEM.2017.8269509
6. R. Berri, F. Osório, A 3D vision system for detecting use of mobile phones while driving. In: International Joint Conference on Neural Networks (IJCNN) (2018), pp. 1–8. https://doi.org/10.1109/IJCNN.2018.8489093
7. Y. Artan, O. Bulan, R.P. Loce, P. Paul, Driver cell phone usage detection from HOV/HOT NIR images. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops (2014), pp. 225–230. https://doi.org/10.1109/CVPRW.2014.42

Hou *et al. J Wireless Com Network*   (2022) 2022:42

Page 11 of 11

8. K. Seshadri, F. Juefei-Xu, D.K. Pal, M. Savvides, C.P. Thor, Driver cell phone usage detection on Strategic Highway Research Program (SHRP2) face view videos. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2015), pp. 35–43. https://doi.org/10.1109/CVPRW.2015.7301397

9. J. -K. Choi, Y. -J. Kwon, K. Kim, J. Jeon, B. Jang, Driver behavior analysis and warning system for digital cockpit based on driving data. In: International Conference on Information and Communication Technology Convergence (ICTC) (2019), pp. 1397–1399. https://doi.org/10.1109/ICTC46691.2019.8939875

10. W. Liu, D. Anguelov, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg, et al., SSD: single shot multibox detector. In: Proceedings of European Conference on Computer Vision (2016), pp. 21–37

11. J. Redmon, S. Divvala, A. Farhadi, R. Girshick, You only look once: unified real-time object detection. In: Proceedings of Computer Society Conference on Computer Vision and Pattern Recognition (2016), pp. 779–788

12. S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1137–1149 (2015)

13. M. Wang, L. Liu, Y. Dong, Research on image neural network detection model based on SSD. J. Tonghua Norm. Univ. (6) (2019)

14. M. Sandler, A. Howard, M. Zhu, et al., MobileNetV2: inverted residuals and linear bottlenecks. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018), pp. 4510–4520

15. A. Howard et al., Searching for MobileNetV3. In: IEEE/CVF International Conference on Computer Vision (ICCV) (2019), pp. 1314–1324. https://doi.org/10.1109/ICCV.2019.00140

16. J. Redmon, A. Farhadi, Yolov3: an incremental improvement. arXiv:1804.02767 (2018)

17. https://aistudio.baidu.com [online]

Publisher's Note