LinDMS: Driver Monitoring System with Embedded System

Sheng-Yu Lin (林聖祐), *Chiou-Shann Fuh* (傅楸善), *Jia-Hao Kang* (康家豪), *Ting-Chi Chang* (張婷淇), *Zhi-Hong He* (何志宏)， *Wei-Cheng You* (游惟丞)

Department of Computer Science and Information Engineering,
National Taiwan University, Taipei, Taiwan,
E-mail: r10945050@ntu.edu.com fuh@csie.ntu.edu.tw

ABSTRACT

The Driver Monitoring System (DMS, Driver Monitoring System) can detect changes in the driver's head, face and eyelid movements through real-time images to alert drivers of their drowsiness or distraction. China and the European Union have stipulated that brand-new vehicles must be equipped with DMS system.

We propose a driver monitoring system based on YOLOv5s (You Only Look Once version 5 small) and use the Kneron KL-520 neural network accelerator chip to run on an embedded device, which can detect driving eyes closed, abnormal head posture, yawning, using mobile phones, smoking, and other behaviors.

Kneron KL-520 is an AI (Artificial Intelligence) chip launched by Kneron in 2019, with 40 nm wafer process technology, a computing power of 350 TOPS, an average power consumption of 300-500 mW, and dynamic model execution. However, there are strict restrictions on the size and structure of machine learning model.

To cope with the limitations of the embedded system, we remove some structures such as the head and neck from the original YOLOv5s, and train the model on the 50,000 images collected by ourselves in the Ubuntu 18.0 environment, and finally deploy machine learning model on the development board equipped with KL-520. The prediction result can be displayed on the display device with an average of 5 FPS (Frames Per Second), which can immediately warn fatigue drivers, and the accuracy of each category can reach more than 90%.

Keywords: *Driver monitoring, CVGIP 2023.*

1. INTRODUCTION

1.1. Driver Monitoring System (DMS)

A driving monitoring system, also known as a driver attention monitor, is a safety system equipped on a vehicle. When the driver is fatigued, distracted, and so on, the system must warn the driver or assist the driver to stop the vehicle. The purpose is to reduce the occurrence of traffic accidents and protect the safety of driving and property.

In 2021, WHO (World Health Organization) had reported that 1.3 million people were killed and 20-50 million people were injured on the roads worldwide [1]. Especially long-distance driving and large vehicles are more likely to cause accidents. Therefore, many countries start to stipulate that brand-new vehicles must be equipped with a driving monitoring system, and a functional driving monitoring system is required.

The driver monitoring system was first proposed by Toyota in 2006 and introduced into its Lexus GS 450h models to monitor the status of the driver and can work together with the collision prevention system in Figure 1. [2].

Fig. 1. The driver monitoring system on the LS 600h, and the red dot is infrared scanner [2].

1.2 Types of DMS

The current DMS can be divided into three categories according to the detection method: based on physiological signals, computer vision, and driving behaviors. Figure 2. shows three categories of DMS and their supporters [10].

DMS based on physiological signals may use ECG (ElectroCardioGraphy), EMG (ElectroMyoGraphy), EEG (ElectroEncephaloGraphy), pulse beat, and breathing frequency as detection targets. The systems based on this method have high sensitivity and high detection speed, however they are expensive.

DMS based on computer vision may use facial feature or eyes as detection targets. For example, PERcent of eye CLOSure (PERCLOS), head position, gaze direction, yawning, micro-sleep, and blink frequency. The systems based on this method have better price–performance ratio.

Finally, DMS based on driver behavior may use steering wheel angle, steering wheel grip, vehicle speed, vehicle offset as detection targets. Although this system may perform poorly compared with the other two, it has a lower cost.

It is hard to say which category is better than the other two, there are products in each category. However, we will focus on computer vision-based DMS in this paper.

Fig. 2. There are mainly three categories of DMS on the market [10].

1.3. Challenges to Computer Vision-Based DMS

In current vision-based monitoring systems, the main visual cues include facial features, hand features, or body features. Sometimes, a visual cue is combined with other cues to make a robust system.

However, vision-based DMS needs to face many challenges: first, variation of image quality. The biggest technical obstacle is its performance under strong or low light, which may lead to completely white (over-saturated) or completely black images (under-saturated) respectively.

Second, variation of drivers: different genders, ages, and races drive cars. Drivers may wear hats, masks, and glasses (including sunglasses), and their heads will appear in various postures during driving. These complex situations will bring great challenges to algorithms and problem definitions.

Third, how to measure fatigue and sleepiness: Every country has its own criteria of DMS, and it is difficult to meet all of the criteria simultaneously.

Fourth, data acquisition and labeling: Computer vision-based measures have high requirements on rich dataset and image quality.

Last but not least, platform computing capabilities: Since DMS needs to warn the driver immediately when a dangerous situation occurs, the speed is very critical. The system needs enough computing power to ensure the video frame runs in real time. However, many current computing processors in embedded system do not have enough computing power, or their power consumption is too high which could cause system overheating.

2. NATIONAL STANDARDS OF DRIVER MONITORING SYSTEM

Since many countries have their own DMS regulations, we are eager to produce a system that complies with most of the current regulations in the world.

2.1. China National Regulation

The Chinese government released the official DMS national standard in 2020, behaviors monitored include closed eyes, yawning, turning head, using mobile phone, and smoking. For these behaviors, there are detailed definitions in Table 1. [3].

Table 1. China National Standard-DMS 2020 [3].

	Behavior	Reminder information request
1	Eyes Closed	Eyes fully closed 2 to 3.5 seconds, the system alerts a warning.
2	Yawning	Mouth aspect ratio greater than 0.6 and holds 3 to 4.5 seconds, the system alerts a warning.
3	Abnormal Head Posture	After head is deflected to the abnormal position and holds 3 to 4.5 seconds, the system alerts a warning.
4	Answering Cell Phone	After cell phone is within 5 cm of face and holds 3 to 4.5 seconds, the system alerts a warning.
5	Smoking	After cigarette is within 2 cm of mouth and holds 2 to 3.5 seconds, the system alerts a warning.

2.2. European Union Regulation

On the other hand, in November 2019, the council of the European Union voted to adopt regulations to mandate the presence of advanced safety systems in automobiles by 2023 [4]. Under the new rules, all motor vehicles including trucks, buses, vans, and sport utility vehicles will have to be equipped with DMS. Functionality must include driver drowsiness and attention warning systems as well as advanced driver distraction warning systems. In the regulations, the levels of driving fatigue are mainly based on KSS (Karolinska Sleepiness Scale) in Figure 3. [5]. KSS is a way to measure the subjective level of sleepiness at a particular time during the day. Generally, the method of measurement is that a recorder would sit next to the driver and ask the driver about his fatigue level every period of time. Regulations adopting KSS as the standard will give manufacturers more flexibility to design their system.

Rating	Verbal descriptions
1	Extremely alert
2	Very alert
3	Alert
4	Fairly alert
5	Neither alert nor sleepy
6	Some signs of sleepiness
7	Sleepy, but no effort to keep alert
8	Sleepy, some effort to keep alert
9	Very sleepy, great effort to keep alert, fighting sleep

Fig. 3. Karolinska Sleepiness Scale [5].

3. METHODOLOGY

3.1. Workflow

In the beginning, we collected the dataset from the Internet and also on our own. After that, we use our personal computer to train a machine learning model and test its functions on a virtual machine where KL520 dongle would be used during the test. Finally, we port the tested model to the Otus8600 development board, which is also equipped with the KL520 chip.

During execution, the computer or development board's camera can receive the driver's facial features and judge whether the diver's state is normal or fatigued. Observed fatigue and distraction characteristics include closed eyes, yawning, abnormal head pose, using a mobile phone, and smoking. When signs of fatigue or distraction appear, the systems warn the driver immediately with a sound in Figure 4. [6].

.

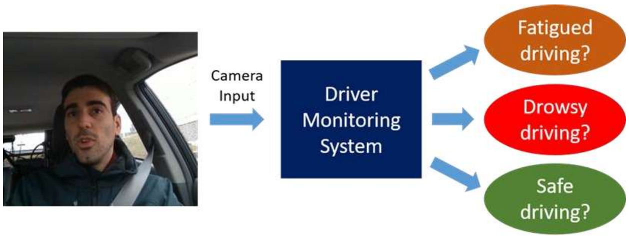


Fig. 4. The system can receive camera input and output the driver's state [6].

3.2. Experiment Setup

We use the following equipment for experiment:

- Personal computer or laptop (Ubuntu18.04);
- RGB (Red, Green, Blue) and IR camera with 1,920×1,080 resolution;
- Kneron KL520 USB AI (Universal Serial Bus Artificial Intelligence) dongle in Figure 5. [7];
- Otus8600 RDK (Reference Design Kit) board in Figure 6.;
- RS-232;
- SD (Secure Digital) card 64GB;

Fig. 4. Kneron KL520 USB Dongle [5].

Fig. 6. Otus8600 RDK board.

The personal computer is for training our own models and testing before porting to embedded systems. RGB and IR cameras are used to collect our own data sets and test experimental results. It should be noted that most DMS systems currently use IR cameras as the main ones, because they are less affected by light, and the comparison is in Figure 7.

Fig. 7. The comparison between RGB camera (left) and IR camera (right) in the night.

KL520 allows experiments to be carried out in different environments as long as it is plugged into any terminal device with a USB (Universal Serial Bus) jack. We will port the model and program to the Otus8600 RDK board after we successfully test it with the KL520 dongle on our own computer. The Otus8600 RDK board also has the KL520 chip. The only difference is that the Otus8600 RDK board lacks a piece of flash memory compared with the dongle.

Finally, the purpose of the SD card is to allow us to update the firmware in the Otus8600 RDK board, and the purpose of RS-232 is to allow the output information of the Otus8600 RDK board to be displayed on our computer.

3.3. Our LinDMS Model

Due to storage space and speed, we choose to use YOLOv5s as our object detection model. In addition, for speed reasons, we also remove the neck of the original YOLOv5, so that the model can have a higher FPS (Frame Per Second) when running. The original neck in YOLO is to up-sample and contact different sizes of feature maps. Removing this part may reduce the prediction ability of the model in multiple scales. Figure 8. shows our LinDMS model.

There are some components in the model, such as SPP (Spatial Pyramid Pooling), CSP (Cross-Stage Partial Connection), and residual block in Figure 9. The principle of SPP is to solve the limitation that the traditional CNN network has to fix the size of the input image by using three different sizes of pooling layers to extract features and then combine them. The use of CSP can reduce the amount of calculation and improve the learning ability of the model. Its method is to make the network have two branches, one part is maintained typically, and the other part is directly concatenated to the output of this block. Finally, the residual block uses skip connections to deepen networks and solve the degradation.

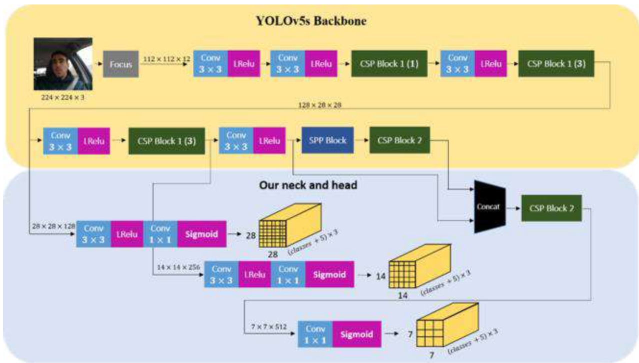


Fig. 8. Our LinDMS model – YOLOv5 reduced.

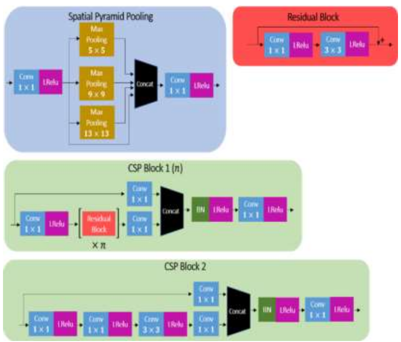


Fig. 9. Components of YOLOv5s.

The loss of DMS model consists of three parts: loss for bounding box, loss for confidence and loss for classification.

- The loss function of LinDMS model can be expressed as:

$$Loss = L_{\text{bbox}} + L_{\text{conf}} + L_{\text{class}}$$

- The loss function for bounding box regression can be expressed as:

$$L_{\text{bbox}} = \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{i,j}^{obj} \left[1 - \text{IoU} + \frac{A^c \setminus U}{A^c} \right]$$

- The loss function for confidence can be expressed as:

$$L_{\text{conf}} = -\sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{i,j}^{obj} \left[C_i^j \log \hat{C}_i^j + \left(1 - C_i^j\right) \log\left(1 - \hat{C}_i^j\right) \right] - \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{i,j}^{noobj} \left[C_i^j \log \hat{C}_i^j + \left(1 - C_i^j\right) \log\left(1 - \hat{C}_i^j\right) \right]$$

- The loss function for classification box regression can be expressed as:

$$L_{\text{class}} = -\sum_{i=0}^{S^2} I_{i,j}^{noobj} \sum_{c \in \text{classes}} \left[P_i(c) \log \hat{P}_i(c) + (1 - P_i(c)) \log\left(1 - \hat{P}_i(c)\right) \right]$$

The loss for the bounding box uses GIoU (Generalized Intersection over Union) instead of IoU (Intersection over Union). GIoU is an improved version of IoU. Even if the bounding boxes do not overlap, the model can be optimized. The comparison between GIoU and IoU is in Figure 10. The loss for confidence and classification use binary cross entropy.

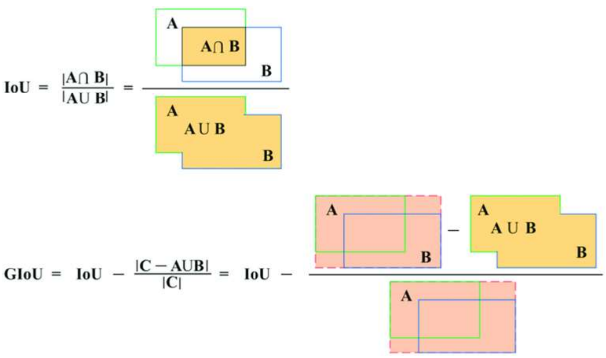


Fig. 10. GIoU and IoU.

3.4. Model Conversion (Pytorch to NEF)

After we have trained our model, we need to go through a process to convert the original Pytorch format into NEF (Neural network Exchange Format) format. NEF is a format that can run on Kneron AI chips in Figure 11. [7].

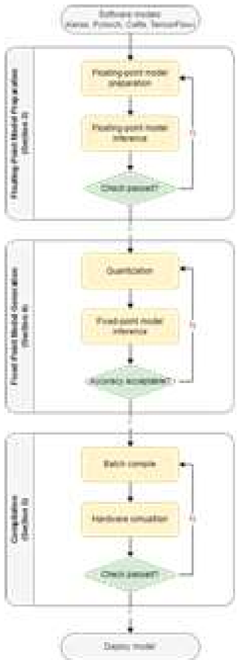


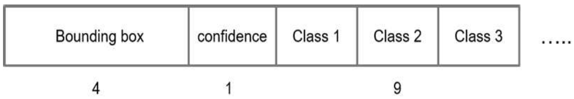
Fig.11.: Conversion from Pytorch to NEF [7].

The working flow of conversion can be divided into three main sections:

1. Floating-point model preparation: At this stage, we will use the ONNX converter provided by Kneron to convert our model into ONNX format and then evaluate the performance of the model and check whether the model has unsupported layers. Then, test the ONNX model and compare the result with the source.
2. Fixed-point model generation: At this stage, we will quantize the model, make the value into a discrete form, and generate a BIE file. We also test the model and compare the result with the previous stage.
3. Compilation: At this stage, we will finally use the batch compiler to convert one or more models into the NEF format and then do hardware simulation to check whether the NEF file can operate as expected.

3.5 Fatigue and Distraction Detection

The final output of our model will be composed of 3 parts, namely the position of the bounding box, the probability of confidence and the probability of each category, shown in below:



1. Position of bounding box: Contains the center point x, y coordinates and width and height of the bounding box.
2. Probability of confidence: The probability that the box contains the object.
3. Probability of each category: Contains nine categories of probability, which are open eyes, closed eyes, yawning, no yawning, normal head pose, abnormal head pose, using mobile phone, and smoking.

Then we will choose the category with the highest probability as the final result. For overlapping boxes, non-maximum suppression is able to ignore the smaller overlapping bounding boxes. However, there may be category exclusion in the output screen, such as open eyes and closed eyes appearing at the same time. If this happens, we will compare the two different bounding boxes and keep the one with the greater class probability. These post-processing processes is in Figure 12.

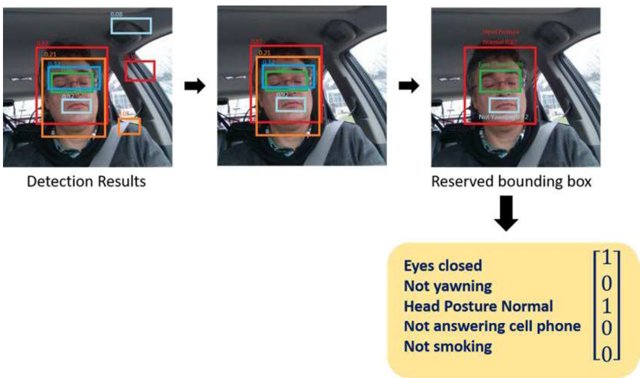


Fig. 12. Post-processing process.

3.5.1. Meeting China DMS Standard

To meet China DMS standard, we designed a debounce function. This function will use the timestamp from the previous few seconds to the present to calculate the frequency of certain dangerous behavior. For example, the law stipulates that the system will issue a warning when the eyes are closed for two seconds, then we will calculate the frequency of closing the eyes from the past 2 seconds to the present, and if this is greater than the threshold, a warning will be issued.

The debounce function can expressed as:

$$threshold = \frac{0.7 \cdot N}{FPS \cdot T}$$

- N: The count of occurrences of the behavior.
- T: Time limit prescribed by standard.
- FPS: Frame Per Second.

3.5.2. Meeting EU DMS Standard

To meet EU regulation, we refer to the method of L. Pauly et al. in 2005 [8]. We simply divide KSS into two parts in Figure 13., normal and drowsy.

Rating	Verbal descriptions	
1	Extremely alert	
2	Very alert	
3	Alert	Normal
4	Fairly alert	
5	Neither alert nor sleepy	
6	Some signs of sleepiness	
7	Sleepy, but no effort to keep alert	
8	Sleepy, some effort to keep alert	Drowsy
9	Very sleepy, great effort to keep alert, fighting sleep	

Fig. 13. KSS can be divided into normal and drowsy [8].

Then we use PERCLOS (Percentage of Eyelid Closure over the pupil over time) to set the threshold for drowsiness.

The formula of PERCLOS can be expressed as:

$$\frac{\text{No of frames in which eyes are closed in one min}}{\text{Total number of frames in one minute}} \times 60$$

The average blink duration of a human being is 100-400 ms, and the number of blinks per minute is 10-15 [9]. From these values, the time interval for which human eyes will be closed in 1 minute would be = 400 x 15 = 6,000 ms for a normal person.

We set a threshold value for PERCLOS to be 6 seconds. If the PERCLOS of someone is over this limit, then the person is considered to be drowsy.

4. DATASET

We collect more than 50,000 images for training and validation, including nine categories. The ratio of training and validation is eight to two, and data augmentation is used on the training dataset.

Our data set mainly comes from three sources, namely YawDD (Yawning Detection Dataset), DMD (Driving Monitoring Dataset), and NTU (National Taiwan University) Dataset. Below we will have a complete introduction to the three sources.

4.1. YawDD: Yawning Detection Dataset

Yawning Detection Dataset is a collection of labeled video data used for developing and evaluating automated yawning detection algorithms. The videos were captured with different camera angles, lighting conditions, and facial expressions. The dataset includes annotations that mark the occurrence of yawns in each video and the corresponding timestamps.

4.2. DMD: Driving Monitoring Dataset

The Driver Monitoring Dataset (DMD) is a collection of labeled video data used for developing and evaluating Driver Monitoring Systems (DMS). The dataset contains videos of drivers captured with an interior-facing camera while performing various driving tasks, such as driving on a highway, on city roads, and in rural areas. The videos were recorded in different lighting conditions, weather conditions, and times of the day.

4.3. NTU Dataset

This is a dataset collected by our lab. The collected images include 20 males and 3 females, with RGB and IR format. Figure 14. shows few examples.

Fig.14. NTU Dataset.

5. EXPERIMENT RESULT

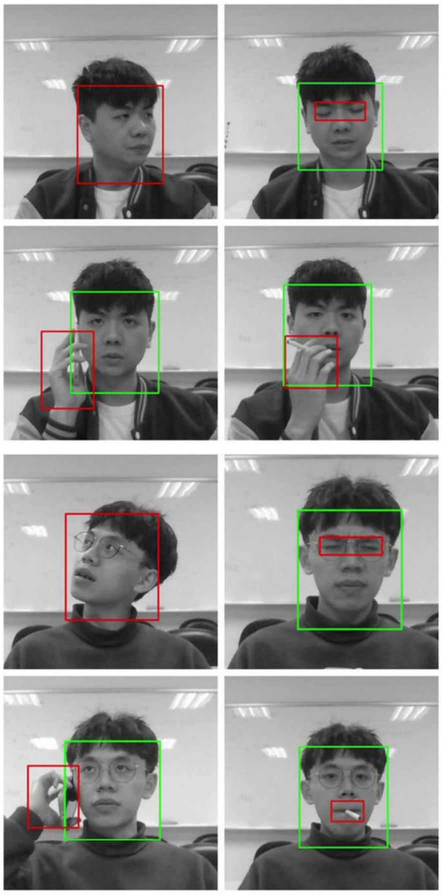
We tested our system. We ask the testers to simulate the driving situation and perform distracted or fatigued behaviors. Five people were in the test, including three men and two women. Each person took 3 to 5 minutes. The resolution of the video was 640x480 pixels; the actual resolution of entering the model 320x320 pixels.

The test results are as follows:

	Accuracy	Precision	Recall
Eyes Closed	0.9738	0.9130	0.9033
Yawning	0.9799	0.8571	0.9056
Head Posture Abnormal	0.9829	0.8437	0.9818
Using A Mobile Phone	0.9301	0.9744	0.8268
Smoking	0.9371	0.9439	0.7426

As shown in the results, the system has a good performance in some categories. However, there is still room for improvement in the recall rate of using mobile phones and smoking. These two categories have more false-negative cases because some gestures of using a mobile phone and smoking a cigarette still cannot be detected. Currently, the reason for detection errors is mostly due to lighting, special gestures, and background effects, and to be improved.

Some successful cases and failed cases are shown below:



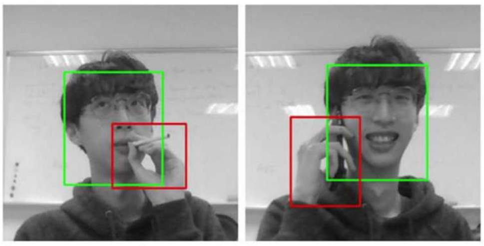


Fig. 15. Successful cases.

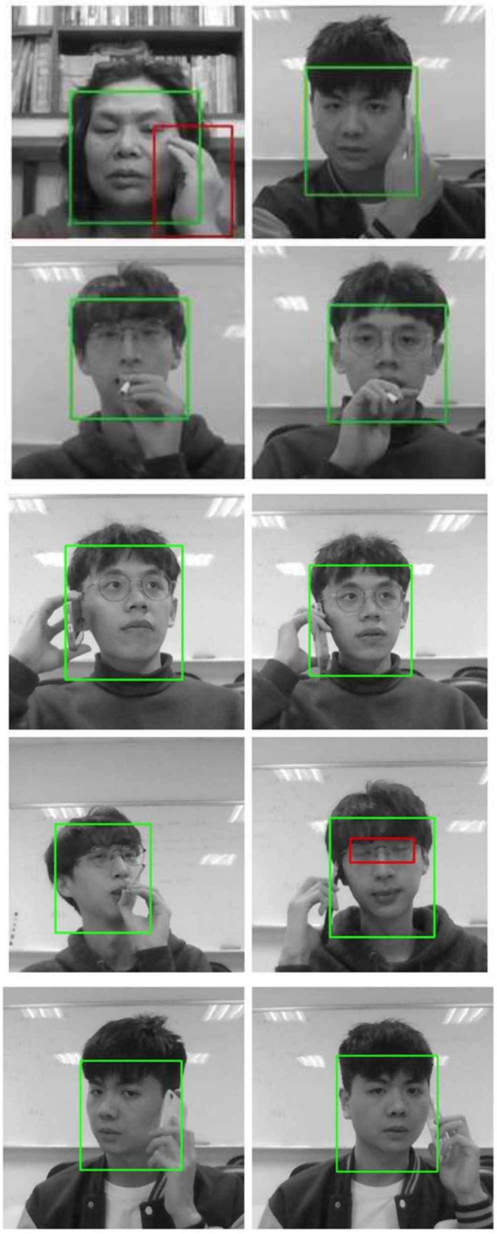


Fig. 16. Failed cases.

6. CONCLUSION

We have developed a DMS (Driver Monitoring System) application based on YOLOv5s that can run on the KL520 embedded device and has the potential for future implementation in vehicles. This application has the advantage of low power consumption and complies with regulations in both China and the European Union. It is capable of detecting driver fatigue and distracted driving. Currently, it achieves a certain level of accuracy in detecting various behaviors. However, further improvements are still needed before practical application.

7. FUTURE WORK

There are still several problems to be solved. The first is occlusions. If the driver wears something such as glasses, hats, and masks or has a light reflection on their face, it may influence the model's accuracy. The second is the issue of race. Since Asians dominate our data set, the model's accuracy for other races has yet to be verified. The third is that the current method needs more evidence to be mapped to KSS. The first two problems may be solved by collecting more data, and the third problem requires more research on fatigue and sleep.

8. REFERENCES

[1] World Health Organization Report, “Road Traffic Injuries,” https://www.who.int/news-room/fact-sheets/detail/road- traffic-injuries, 2021.

[2] Wikipedia, “Driver Monitoring System,” https://en.wikipedia.org/wiki/Driver_monitoring_system, 2022.

[3] China National Standard (中华人民共和国国家标准), “Performance Requirements and Test Methods for Driver Attention Monitoring System,” http://www.catarc.org.cn/upload/202012/29/202012290851061182.pdf, 2020.

[4] European Commission, “Commission Delegated Regulation (EU),” https://members.wto.org/crnattachments/2021/TBT/EEC/21_0621_00_e.pdf, 2021.

[5] T. Åkerstedt, “Karolinska Sleepiness Scale,” https://www.med.upenn.edu/cbti/assets/user-content/documents/Karolinska%20Sleepiness%20Scale%20(KSS)%20Chapter.pdf, 1990.

[6] B. C. Kung, “Driver Monitoring System with Embedded Artificial Intelligence,” Master Thesis, Department of Computer Science and Information Engineering, National Taiwan University, 2022.

[7] Kneron, “Kneron AI Documentations,” https://doc.kneron.com/docs/#, 2019.

[8] L. Pauly and D. Sankar, “Detection of Drowsiness Based on HOG Features and SVM Classifiers,” *Proceedings of IEEE International Conference on Research in Computational Intelligence and Communication Networks*, Kolkata, India, doi: 10.1109/ICRCICN.2015.7434232, pp. 181-186, 2015.

[9] Wikipedia, “Blinking,” https://en.wikipedia.org/wiki/Blinking., 2023.

[10] Datik Irizar Group, “An Industry Review on Driver Fatigue Systems for Fleet Operators,” https://www.datik.es/en/datik-3/news/an-industry-review-on-driver-fatigue-systems-for-fleet-operators-2, 2018.