Fuzzy Systems and Data Mining IX
A.J. Tallón-Ballesteros and R. Beltrán-Barba (Eds.)

doi:10.3233/FAIA231097

Intelligent Driver Behavior and Emotion Detection System

Chi-Yat Lau [a], Man-Ching Yuen [b,1], Chi-Wai Yung [b], Ka-Ho Yeung [a], Ho-Yin Chung [a], Ho-Wang Ngai [a], Chau-Wai Wong [a], Ching-Yee Wong [a], Alan Li [c]
[a] *Department of Information Technology, Vocational Training Council, China*
[b] *Department of Applied Data Science, Hong Kong Shue Yan University, China*
[c] *Sampras (HK) Limited, China*

Abstract. According to the road traffic injuries fact sheet from the World Health Organization, approximately 1.3 million people die each year because of road traffic crashes. There are many risk factors and one of the most important is distracted driving, especially caused by using a mobile phone. The fact sheet shows that drivers using mobile phones are approximately 4 times more likely to be involved in a crash than drivers not using a mobile phone. Apart from this, dangerous driving behaviors and driver emotions are growing concerns for road safety. This project aims to develop a model for detecting dangerous driving behaviors and analyzing drivers' emotions in order to avoid and minimize traffic accidents. We focus on detecting several distractive behaviors, bad driving practices, and abnormal emotions. We carry out experiments to evaluate the system performance through the integration of our system with existing infrastructure. To the best of our knowledge, we design and develop the first small and portable device to collect data from drivers, then the collected data is transferred to the cloud for further analysis. The trial service is already available for drivers for a few months in Hong Kong.

Keywords. Driver behavior, emotion detection, driving, traffic accidents

1. Introduction

Since traffic accidents, casualties and vehicle licenses are keep increasing, the main reason of accidents is distracted driving. In the literature, there are many traffic monitoring systems [1], but the number of traffic accidents is still high. Moreover, the highest accident vehicle involvements are private cars. Therefore, we design an Intelligent Driver Behavior and Emotion Detection system for monitoring driver behavior which is for private car usage.

"CAOME", our project is proposed on faith to protect road users' safety by reducing risky driving practices. Other than some uncontrollable and unexpected situations caused by passengers and pedestrians, drivers are inevitably to take responsibility for any accidents. Hence, our project's target users are drivers.

This research project is composed of four parts, which include model development, cloud service building, IoT deployment, and user platform development. We have

[1] Corresponding Author, Man-Ching YUEN, Department of Applied Data Science. Hong Kong Shue Yan University, Hong Kong, China; E-mail: mcyuen@hksyu.edu.

applied several emerging technologies in the whole project, which include Artificial Intelligence (AI), Machine Learning, Deep Learning, and 5G.

First, our project developed an intelligent driving system with a customized detective model for detecting drivers' dangerous behaviors including hands-off driving, smoking, using the mobile phone, and physical reactions of tiredness. Moreover, we have built an analysis model to capture the driver's facial expressions and then identify the driver's emotions of anger, fear, happiness, sadness, and neutrality. We design and develop the sensors that we used in the project.

Second, we build and develop our service on Alibaba Cloud. The system operation and maintenance are performed on the cloud server. We build connections and communication between the cloud and IoT devices.

Third, we make our custom device to ensure the best quality for I/O. We choose the most suitable parts for handling nighttime driving and offline operation.

Finally, we provide user-friendly platforms on the website and smartphone app. Users can obtain our service through easy operations.

The organization of this paper is as follows. Section 2 describes motivation. Section 3 presents our data analytical framework. Section 4 shows the object detection model and performance evaluation. Section 5 presents the system function. Section 6 shows the solution implementation. Section 7 draws out the conclusion.

2. Motivation

2.1. Comparison of major driver contributory factors of traffic accident

Refer to Figure 1. Corresponding to comparison of major driver contributory factors of traffic accidents from Hong Kong Police Traffic Report [2], traffic accidents caused by inattentive driving are in first place in the past 3 years and 2021 have proliferated obviously.

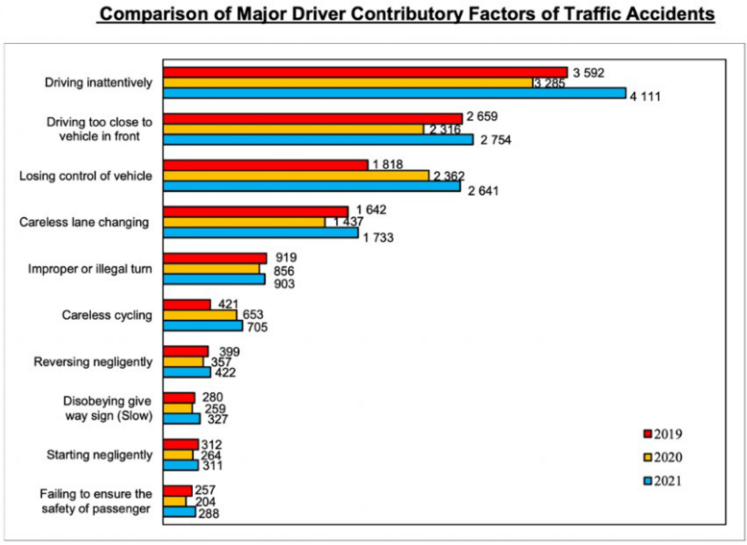

Figure 1. Comparison of major driver contributory factors of traffic accident from Hong Kong Police [2]

2.2. Our Contribution

Concerning the problem issued above, we developed a customized system, CAOME, for monitoring drivers' driving behaviors. Our project's ultimate goal is to minimize traffic accidents caused by drivers. The contributions of our system are:

1. Achieve stable performance in daytime and nighttime driving.
2. Develop automated and fully connected operations between clients and servers.
3. Build a self-learning and improvement model for future optimization.

3. Data Analytical Framework

3.1. Data acquisition and understanding

Since we have two models for driver behavior detection and emotion analysis, we have created two corresponding datasets.

3.1.1 Dataset for driver behavior detection

We have some of the functions performed through this model, which include hands-off from steering wheels, smoking, and using the mobile phone. Thus, we have collected images of cigarettes, mobile phones, steering wheels with hands, and steering wheels without hands. We looked for images that are under driving conditions. These images are collected from different channels, such as Kaggle, Google search engine, and YouTube videos. There were around 18,000 images collected.

3.1.2 Dataset for driver emotion analysis

In the emotion analysis model, we classify the driver's emotions into 5 categories, **angry, fear, happy, sad, and neutral**. We have collected 54,796 images from different open-source datasets, such as Kaggle, RAF-ML dataset, Basel Face dataset, Chicago Face dataset, and Color FERET dataset. We classify into five emotion categories which are neutral, happy, angry, sad, and fear. There are criteria regarding data acquisition;

1. Clear detection object.
2. High resolution.
3. Light distributed evenly as much as possible.

We found that there exist problematic data due to blurred objects, too dark for object detection, not related objects, low resolution, images generated from video recorded at low light environments. All problematic data in these datasets is removed.

3.2. Data Preprocessing

Data collected from various channels which are Kaggle, Search engine, and YouTube. Around 23,00 pictures collected for the driver behavior model, 54,796 pictures collected for the emotion model.

All image data is checked manually if there is any irregularity which is mentioned before. After data validation, data was labeled according to the image features. Data is labeled by labelstudio which exports TXT format, and YOLOv5 accepts TXT format. Therefore, no data format converters are needed.

4. Object Detection Model and Performance Evaluation

Object detection is a popular computer vision task that involves identifying and localizing objects of interest within an image or video frame. It has numerous applications in fields such as surveillance, autonomous driving, robotics, and augmented reality. Object detection models are designed to automatically identify and locate objects within an image or video. There are several types of object detection models [3, 4, 5, 6]. Object detection models are poised to play an increasingly important role in many real-world applications [7].

4.1. Algorithm of YOLO

In this project, we use an object detection model to detect if the driver is in safety status, smoking, hands-off or using a phone. YOLO algorithm [5] performs real-time object detection using CNN [3, 4]. YOLO is well-fit for real-time applications like video surveillance and autonomous driving due to its effective architecture and capacity to recognize objects at various scales [8].

The commonality among all object detection architectures is that, as shown in Figure 2 [9], the input picture characteristics are first compressed by the feature extractor (Backbone) before being sent to the object detector (which includes the Detection Neck and Detection Head). In order to prepare for the detection stage in Detection Head (or Head), Detection Neck (or Neck) functions as a feature aggregation that is tasked with mixing and combining the features created in the Backbone.

Here, the distinction is that the head is in charge of each bounding box's detection, localization, and classification. As shown in Figure 2, the one-stage detector implements the two jobs simultaneously (Dense Detection), whereas the two tasks are implemented separately by the two-stage detector and combined subsequently [9].

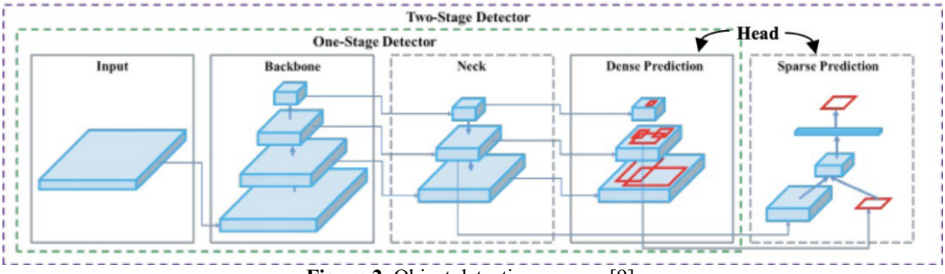

Figure 2. Object detection process [9].

4.2. Performance Evaluation Measures

To evaluate the speed of recognition and accuracy of our model, we use the following performance measures: IoU, precision, recall and mAP.

Precision is the ratio of positive prediction value over all predictions. The formula of precision is (Eq. 1). Recall is the ratio of positive prediction value over ground truth. The formula of recall is (Eq. 2). Eq. 3 shows the calculation of F1 score. Intersection Over Union (IoU) is the measure used to calculate the overlap of the area where the predicted bounding box and the actual bounding box intersect. The formula of IoU is described at (Eq. 4). Mean average precision (mAP) refers to the mean average of the Average Precision (AP) values for all classes. Average Precision is the average precision of all predictions.

$$Precision = \frac{TP}{TP+FP} \quad \text{(Eq. 1).}$$

$$Recall = \frac{TP}{TP+FN} \quad \text{(Eq. 2).}$$

$$F1\ Score = 2 \times \frac{recall \times precision}{recall + precision} \quad \text{(Eq. 3).}$$

$$IoU = \frac{R_A \cap R_P}{R_A \cup R_P} \quad \text{(Eq. 4).}$$

4.3. Performance Evaluation on Object Detection Model

All of the models' results are presented in Table 1. As can be seen, both YOLOv5 models achieved the best mAP value, which is 99.5. Other models' mAP levels are higher. The performance of the YOLOv4 small model outperformed the competition in terms of training speed. These findings show that while the YOLOv4 small model's training duration is the best, its mAP value is the worst.

There are fewer convolutional network layers in little models than in larger ones. Although YOLOv5s requires twice as much training time as the competition in tiny models, it has surpassed YOLOv4 and YOLOv3 tiny.

Table 1. Performance results of different models

Result	mAP@0.5 IoU	loss	Precision	Recall	Training Time (mins)
Yolov3 tiny	90.3	0.74	0.93	0.83	55
Yolov3	92.9	0.18	0.96	0.93	297
Yolov4 tiny	86.2	0.09	0.96	0.79	**48**
Yolov4	96.6	1.35	0.95	0.97	379
Yolov5 s	**99.5**	0.01	0.997	**0.99853**	108
Yolov5 xl	**99.5**	**0.007**	**1.00**	0.99783	279

In summary, we find out that YOLOv5 has the most consistent and superior performance among all these models. Therefore, YOLOv5 was selected even though its training period was longer than that of other models since we wanted a model that would be very accurate given the situation.

4.4. Performance Evaluation on Emotion Detection Model

For emotion detection, it uses EfficientNetV2s. To determine the model's ideal parameters, we apply the finetune method. The knowledge and representations that a pre-trained model has gained from a big dataset are used in fine-tuning. In comparison to training from scratch, the fine-tuning procedure requires fewer iterations and less training time when the model is started using these pre-trained weights. And EffcientNetV2s provides faster training times and highest accuracy compared to the previous version.

We thoroughly explored model scaling and discovered that performance may be enhanced by adjusting network depth, width, and resolution. A neural convolutional network that is more accurate and efficient than previously, so it makes a lot of progress. EfficientNetV2 obtained the best top-1 accuracy of 84.3% on ImageNet while downscaling by a factor of 8.4 and performing 6.1x faster inference than the best Convolutional neural network [10].

5. System Function

The model contains two separate models which are driver behavior detection and emotion classification.

5.1. Driver behavior detection

In this model, face, hand and eyes landmarks are used for analysis to detect driver behavior. Therefore, Face Mesh and Hand Mesh from Mediapipe packages and Dlib package are used for extracting user face, hand and eyes landmarks respectively. All functions are performed in “model.py”. The system will issue a warning once the following driver’s behavior is detected:

- Holding phone
- Looking around
- Eye closed
- Yawning
- Camera covered
- Smoking

5.2. Driver emotion classification

“Emotion” function is to perform a driver’s emotion classification. There are a total five classes of the classification, which are neutral, happy, sad, fear, and angry. It predicts emotions from a given image using face detection and emotion recognition models.

6. Solution Implementation

6.1. System architecture

Figure 3 shows the system architecture. Our system adopts a three-tiered system architecture, which divides the application into physical and logical layers. This architecture allows for better scalability, maintainability, and reliability. Each layer has different responsibilities and functions.

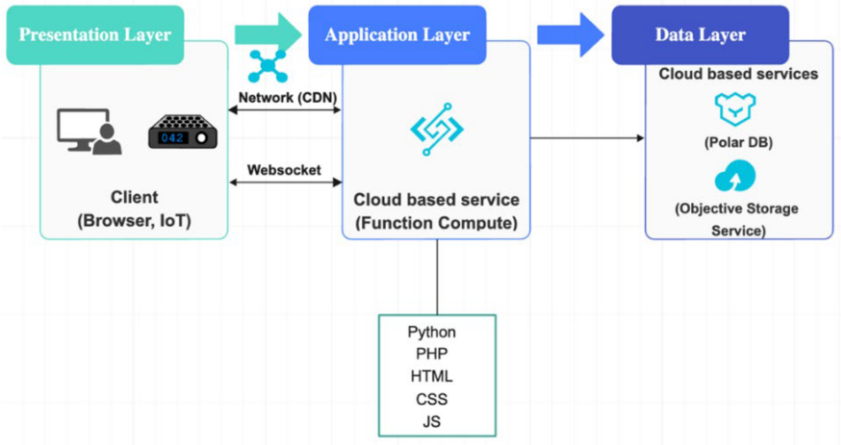

Figure 3. System architecture

6.2. System Flow

Figure 4 shows the system flow. The end user can start the operation through either the IoT device or our website. As soon as the camera starts streaming, the images are captured and collected. These images are then uploaded to the cloud-based server via web-socket. The processed image and alert sounds are retrieved through web-socket and sent back to the device. Figure 5 shows our system website.

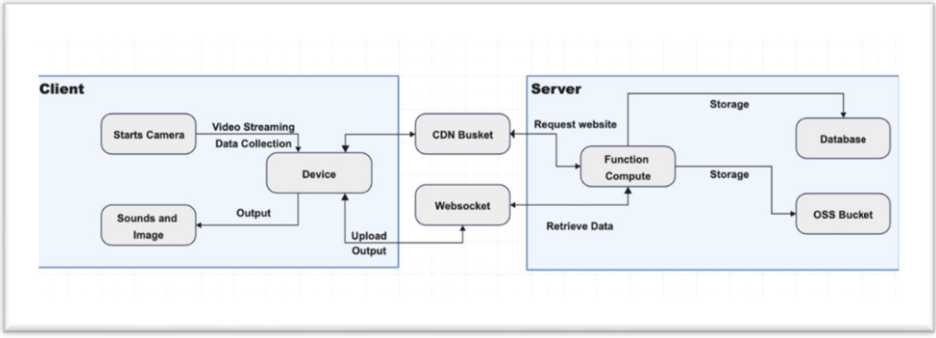

Figure 4. System flow

6.3. Website

Figure 5 shows our system website. Our one-page website features a user-friendly interface that provides a simple way to view the system's performance. Users can start and stop the streaming by clicking buttons, and the real-time streaming is displayed in the center of the page. The images are processed and retrieved with model performance results and alert sounds.

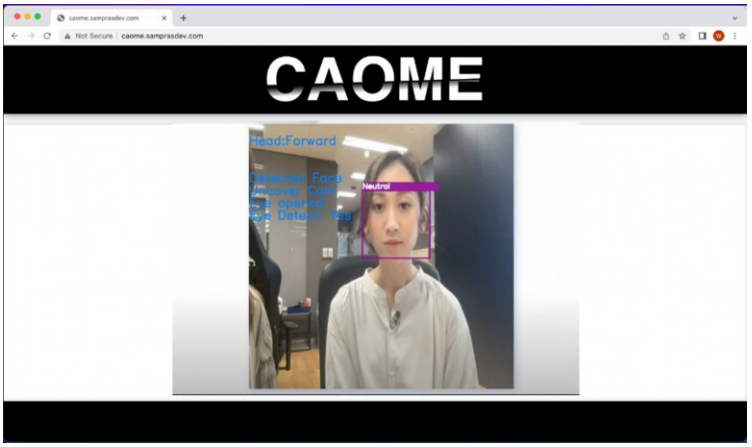

Figure 5. Our System Website

7. Conclusion

In this project, we have successfully developed a comprehensive driver behavior and emotion detection model. Our model ensures accurate detection and analysis of various driver-related factors, including phone usage, smoking, and instances of eyes-closed scenarios. We have applied data preparation techniques, machine learning algorithms, cloud services, and Vehicle-to-Everything (V2X) technology to build a robust driver behavior and emotion detection system.

By capturing real-time frames and uploading them to the cloud for computation and analysis, our system can promptly alert drivers when it detects inappropriate driving behavior. The primary objective of our model is to minimize traffic accidents caused by such behaviors. By effectively identifying and monitoring driver actions, our system helps to enhance the overall road safety and prevent the potential risks associated with distracted driving.

Our system aims to raise safety awareness among drivers by providing real-time alerts and feedback on their behaviors. By actively monitoring and analyzing driver actions, we not only strive to prevent traffic accidents but also protect other drivers and pedestrians on the road. Through the integration of our system with existing infrastructure, we can call for the smart city and reduce the burden on police forces in handling traffic accidents, allowing them to allocate resources more efficiently.

Acknowledgements

This research was in part supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. UGC/FDS15/E02/20 and UGC/FDS15/E01/21).

References

[1] Lau C. Y., Yuen M. C., Yeung K. H., Fan C. P., Ko O. Y., Ngan L. W., Tam W. C., Yeung W. N., "PC-based Intelligent Traffic Monitoring System with Real-time Analysis for Smart Cities", Proceedings of 2022 14th International Conference on Communication Systems & Networks (COMSNETS), IEEE, January 2022.
[2] Hong Kong Police Traffic Report https://www.police.gov.hk/ppp_en/09_statistics/ts.html
[3] Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. Advances in neural information processing systems, 91-99.
[4] He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017). Mask R-CNN. Proceedings of the IEEE international conference on computer vision, 2961-2969.
[5] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE conference on computer vision and pattern recognition, 779-788.
[6] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., & Reed, S. (2016). SSD: Single Shot MultiBox Detector. European conference on computer vision, 21-37.
[7] Han, C., Ye, Y., & Zhong, G. (2021). Object Detection in 20 Years: A Survey. arXiv preprint arXiv:2105.01988.
[8] Jiang, P. et al. (2022) A review of Yolo algorithm developments, Procedia Computer Science. Available at: https://www.sciencedirect.com/science/article/pii/S1877050922001363 [Accessed: 10 May 2023].
[9] Solawetz, J. (2020). YOLOv5 New Version - Improvements And Evaluation. Roboflow. Search date 04.2020. https://blog.roboflow.com/yolov5-improvements-and-evaluation/
[10] Koonce, B. (2021). EfficientNet. In: Convolutional Neural Networks with Swift for Tensorflow. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-6168-2_10