YOLO-Based Deep Learning Design for In-Cabin Monitoring System with Fisheye-Lens Camera

Yen-Sok Poon, Chih-Chun Lin, Yu-Hsuan Liu, and Chih-Peng Fan*
Department of Electrical Engineering, National Chung Hsing University, Taichung city, Taiwan (R.O.C.)
Email: a8012183@gmail.com, karta3437002@gmail.com, 890706123jason@gmail.com, cpfan@dragon.nchu.edu.tw*

2022 IEEE International Conference on Consumer Electronics (ICCE) | 978-1-6654-4154-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCE53296.2022.9730235

***Abstract*— To exploit an image-based in-cabin monitoring system for driving behavior and in-vehicle occupants detections to improve driving safety, in this paper, by installing a fisheye-lens camera at the in-car roof center and by using RGB-format images as inputs, the YOLO-based deep learning models, including YOLOv3-tiny, YOLOv3-tiny-3l, YOLO-fastest, YOLO-fastest-xl, and YOLO-fastest-three scales, are studied to be candidate detectors. The proposed in-cabin monitoring design can detect the normal and distracted driving cases and in-vehicle occupants including back seat passengers and pet dogs. The experimental results show that the YOLO-fastest-three scales model performs the best metrics for F1-Score and mAP, which are 95.89% and 97.16%, respectively. The YOLO-fastest-xl model has the best metric for false negative rate (FNR), which is 2.63%. By the software realization, the proposed design executes up to 30 frames per second (FPS) with the GPU-based embedded device.**

Keywords*— in-cabin monitoring, deep learning, YOLO, fisheye-lens camera

I. INTRODUCTION

Lately, many efforts have been made to meet the demands of advanced vehicle automation, and the in-cabin monitoring function of vehicles played an important [1]. It guaranteed both the driver and the passenger a safe and comfortable trip. The in-cabin driver's behavior recognition technology was broadly divided into two categories: the "contact" and "non-contact" designs [1]. One of the contact techniques used to detect driver distraction behavior was based on the electroencephalogram (EEG) signals to measure brain activity. But for the contact design, the system required the contact patch between the device and the driver to obtain driver's measurement data. In [2], detecting passenger discomfort required further integration and synchronization with the latest built-in warning system that reported to the emergency response team to rescue passengers. An automated real-time system in the seat was designed and developed to detect physical complaints of passengers while driving [2]. On the other hand, the non-contact driving behavior recognition technology did not directly contact with the driver while it took into account driver comfort and ease of use. The system performed digital image processing of the captured image, and then analyzed and inferred the driver's behavior.

Recent advances in artificial intelligence (AI) have enabled a multitude of new applications and support systems to solve automated problems for the in-cabin monitoring design. By using AI techniques for the in-cabin driving safety and driving comfort, the AI technology has a promising future in the sense of autonomous driving. In [3], the design detected the driving drowsiness by using deep learning and computer vision technologies. When the driver was distracted, the warning signal sounded. The proposed CNN model was applied to detect the position of the face and eyes for the driver. The driving behavior detection design was divided into three functions, which involved the eye predict, face predict, and hand predict. The design in [4] proposed a driver's mobile phone detection based on deep learning technologies. Firstly, the progressive calibration network (PCN) was used to detect the face area for tracking driver's face. Next, the CNN-based driver mobile phone detection method was utilized to detect mobile phones when the driver used the mobile phone in driving. The work in [5] used AI video comprehension technology to monitor vehicle occupants. Fully autonomous vehicles did not have drivers, only passengers. For pay-as-you-go driving services, there was no car owner, so no one was responsible. Passengers were unknown or unfamiliar with driver services and public transport. In all of these use cases, it was important to ensure that all occupants in the vehicle were safe and protected, and that the vehicle was protected from improper uses by occupants and external threats. In-vehicle monitoring was required due to the growing demand for "drive for rent" and "car for rent" models, and the in-cabin vehicle safety about the future adoption of autonomous robot taxis and robot buses became important. It was important to monitor the driver's attention, position, and movement in real time. In [6], the potential application area of Time of Flight (ToF) went far beyond this application, and the 3D ToF cameras provided the driver and in-cabin surveillance with a single wide-angle camera (e.g. 110 degrees).

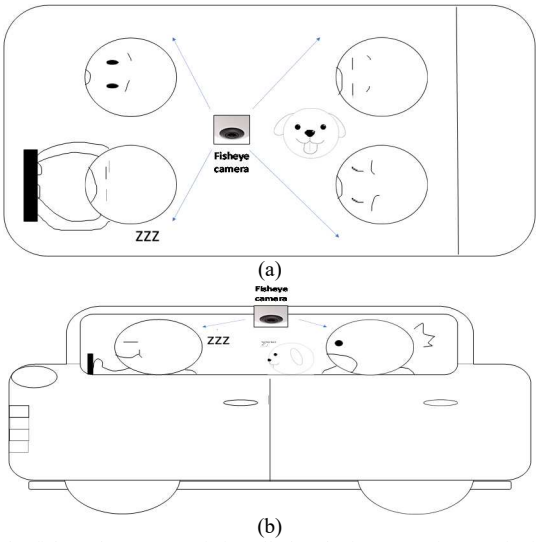


Fig. 1 The fisheye lens camera is installed at the in-car roof center for in-cabin monitoring system: (a) Top view, and (b) Side view

In [7], if there were no human drivers in a fully autonomous vehicle (FAV), the transport system always monitored the interior of the vehicles. Adequate monitoring in the vehicle was not enough for FAV, it was essential to ensure the safety of people and vehicles. A single AI-based camera was proposed to monitor and understand occupants and their behavior in the cabin. In addition, the requirements for interior surveillance were presented, highlighting various irregular situations were taken into account to enable the occupants for driving safely. Besides, the availability of such use cases and data sets was so low that the study in [7] developed their own database.

In this study, our motivations are : The in-cabin monitoring functions can be achieved by using one top camera and deep-learning based technologies. When the driver is sleepy and closes his/her eyes, the system cannot detect the facial status directly by using only one top camera; however, to obtain the drowsy status of the driver, we can use the additional sensor to detect the driving operation status which can represent the changes in driving behavior. For example, steering wheel, acceleration, braking, shifting, etc. Besides, we can also apply the extra sensor to detect the vehicle behavior status which describes the changes in vehicle behavior. For example, speed, lane departures, rapid corners, snakes, etc.

By using a fisheye camera at the in-car roof center which is shown in Fig. 1, the YOLO based models are trained to detect the in-cabin driver behavior and in-vehicle occupants in this paper. The YOLO based models can accurately recognize the driver's distracted behavior, and the passengers and pet dogs are also detected effectively, and then the driving safety will be improved by the proposed design.

II. The Used YOLO-based Models

The YOLO series [9, 10] utilized a multi-scale input training scheme for the YOLOv2 model. Simultaneously, the application of multi-scale training strategy made the model adapt to different image sizes, and it raised the robustness of the YOLO model. For the YOLOv3 model [11], the difference from the previous two versions was that the original single-label multi-class "Softmax" designed for classification was transformed to the multi-label multi-class logistic. The design purpose was to modify that if an object belonged to multiple categories, the Softmax function could not work properly. Besides, the logistic regression layer by the "Sigmoid" function could constrain the range of inputs. YOLOv3-tiny was a light version of YOLOv3. Although the accuracy of the YOLOv3-tiny model was slightly reduced, the model was designed for the real-time detection with an embedded platform, and YOLOv3-tiny was appropriate to be as the candidate.

Next, YOLOv3-tiny-3l was an extended edition of the YOLOv3-tiny model. For high accuracy requirements, when the selected datasets had more small objects, the YOLOv3-tiny-3l model performed better than the YOLOv3-tiny model. The YOLO-fastest [12, 13] model was the modified edition in YOLO series, and the YOLO-fastest model was the lightest and fastest design used for object detections. The YOLO-fastest model performed well, and the numbers of required parameters and calculations were also small. When the input size is 416x416 pixels, the final output of the three feature maps will be 13x13, 26x26, and 52x52. To keep the number of anchor boxes, the original 6 anchor boxes will be changed, and the number of anchor boxes will be adapted to 9. The architecture is selected to be as the proposed YOLO fastest-three scales model. Fig. 2 depicts some parts of the final-stage architecture of the YOLO-fastest-three scales model, which is visualized by Netron [8]. The YOLO-fastest-xl model was an extended edition of the YOLO-fastest model, and YOLO-fastest-xl doubled the number of filters in each layer of YOLO-fastest to enhance the accuracy of detections.

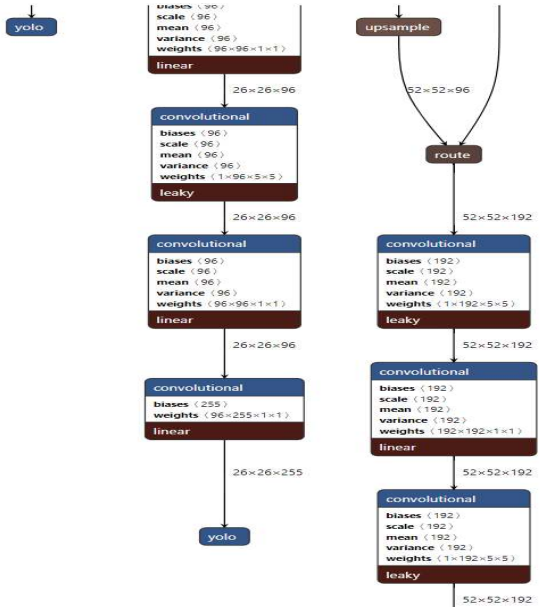


Fig. 2 Parts of the final-stage architecture of the YOLO-fastest-three scales model visualized by Netron

III. Proposed In-Cabin Monitoring Design

For the proposed in-cabin monitoring system, Fig. 3 depicts the proposed design flow. In this work, the self-collected datasets shown in Fig. 4 are produced by the in-car fisheye-lens camera. The sensor of the used camera provides the most effective resolution with 3264x2448 pixels. The field of view (FOV) of the used camera is 360 degrees, and the horizontal view angle is 210 degrees. Besides, the frame rates of the used fisheye camera is 30fps. Before preparing the training datasets to recognize the in-cabin driving behaviors and in-vehicle occupants, the driving behaviors and the in-vehicle occupants are defined by the four corresponding categories. Table 1 defines the four in-cabin monitoring cases for this study. To produce the necessary annotation file for the in-cabin monitoring design, the auxiliary software tool, called by "LabelImg", is utilized to provide the annotations by selecting the bounding box of object. After the labelling process, the collected in-cabin datasets are divided into three parts for the training, validation, and testing processes. Table 2 lists the number of images used for the training, validation, and testing processes. By using data augmentation process [14], more in-cabin images are generated to be used for training the YOLO-based model, and the overfitting problem, which causes by insufficient training images, can be conquered. Table 3 describes the parameters setting used for the applied data augmentation process. To generate the YOLO-based models by the framework, i.e. Darknet [15], the corresponding configuration file must be

created, and various structures of layers for YOLO are generated based on the file. In addition to the parameters of batch, subdivisions, width, height, channels, momentum, decay, and learning rate, the layer's structures for convolution, max pooling, and YOLO are also required to construct the corresponding YOLO models. Besides, at the YOLO layer, the number, length, and width of the anchor box are computed by using the K-means methodology.

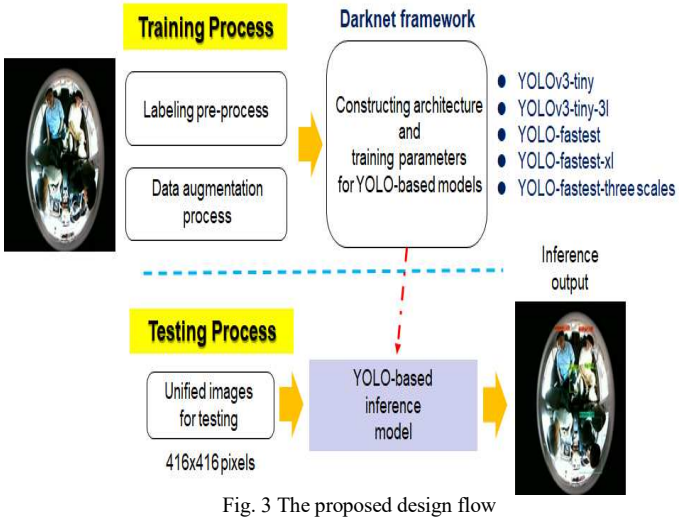


Fig. 3 The proposed design flow

Fig. 4 Self-collected datasets by using the in-car fisheye-lens camera

Table 1 Definitions of the four in-cabin monitoring cases

Four cases	Definitions
Normal driving	The driver holds the steering wheel with his head facing forward
Passenger	Objects of back-seat passengers
Dog	Objects of pet dogs
Distracted driving	The driver's hands leave the steering wheel, and the driver's head is not facing forward, or the driver is engaged in other tasks

Table 2 Number of images for the training, validation, and testing modes

Four categories	Training	Validation	Testing
Normal driving	302	45	36
Passenger	850	109	107
Dog	810	110	96
Distracted driving	533	59	65

Table 4 lists the parameters used for training the applied YOLO-based models. Moreover, the input size of the YOLO-based models is unified as 416x416 pixels, and the number of iterations is 40,200 used for the training process.

Table 3 Parameters setting for the data augmentation process

Parameters	Values
Angle	0
Saturation	1.5
Exposure	1.5
Hue	0.1
Jitter	0.3
Random	1
Mosaic	1

Table 4 Parameters for training the applied YOLO-based models

Width x Height	416x416 pixels
Channels	3
Momentum	0.9
Decay	0.0005
Learning rate	0.001
Learning rate policy	steps
Steps	[16000, 18000]
Scales	[0.1, 0.1]
Max iteration	20200
GPU	PC2080

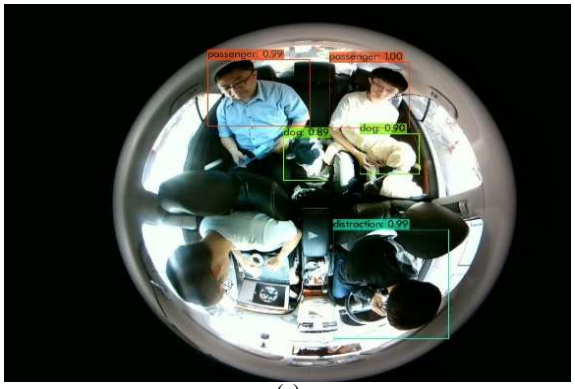
IV. Experimental Results and Comparisons

To evaluate the utilized deep learning models, the evaluation indices, including F1-Score, FNR (False Negative Rate), and mAP (mean Average Precision), are applied for this work. The labeled in-cabin images are separated into the training, validation, and testing sets as the ratios of 8:1:1, respectively, and the five different YOLO-based deep learning models are used for the training, verification, testing, and inference processes. Table 5 lists the performance comparison of F1-Score, FNR, and mAP among the five YOLO-based models. Based on the same parameter settings, the YOLO-fastest-three scales performs the best metrics for F1-Score and mAP, which are 95.89% and 97.16%, respectively, and the YOLO-fastest-xl has the best metric for FNR, which is 2.63%. Figs. 5 and 6 demonstrates the detection results by the proposed YOLO-based in-cabin monitoring design. Because the structures of YOLO-based models are relatively thin and light, and the applied YOLO models supports the real-time computations with the GPU-based platforms. For embedded software implementation with NVIDIA Xavier device, the proposed design performs up to 30 frames per second for the developed in-cabin monitoring applications.

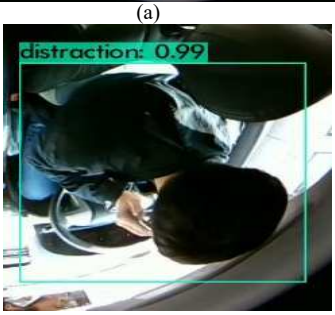
(a)

(b)

Fig. 5 The detection results by the proposed design:
(a) Two back-seat passengers (one wears mask), two dogs, and the driver
(b) Zoom-in diagram of a normal driving case



(a)



(b)

Fig. 6 The detection results by the proposed design:
(a) Two back-seat passengers without wearing mask, two dogs, and the driver
(b) Zoom-in diagram of a distracted driving case when the driver uses a mobile phone

Table 5 Comparison results of the five applied YOLO-based models

YOLO-based Models	F1-Score	FNR	mAP
YOLOv3-tiny	95.78%	5.48%	96.84%
YOLOv3-tiny-3l	95.88%	3.07%	96.78%
YOLO-fastest	94.42%	3.51%	96.42%
YOLO-fastest-xl	95.48%	**2.63%**	96.64%
YOLO-fastest-three scales	**95.89%**	2.85%	**97.16%**

V. Conclusions

To improve driving safety, an image-based in-cabin monitoring system is proposed for the driving behavior and in-vehicle occupants detections. By using a RGB-format fisheye camera located at the in-car roof center, the five YOLO-based deep learning models are selected to be the detectors. The proposed in-cabin monitoring design recognizes the normal driving condition, distracted driving condition, and in-vehicle occupants, which include back-seat passengers and pet dogs. The experimental results reveals that the YOLO-fastest-three scales model performs the best metrics for F1-Score and mAP, which are 95.89% and 97.16%, respectively, and the YOLO-fastest-xl model has the best metric for FNR, which is 2.63%. The proposed design executes up to 30 FPS by the NVIDIA Xavier platform. In future works, the next-generation YOLO series or other CNN structures will be chosen for the further studies. Moreover, the training datasets will be enriched, and the in-cabin monitoring functions will be expanded by adding more categories and behaviors.

Acknowledgment

This work was financially supported by the Ministry of Science and Technology (MOST) under Grant No. 109-2218-E-005-008.

References

[1] Y. Rong, C. Han, C. Hellert, A. Loyal, and E. Kasneci, “Artificial Intelligence Methods in In-Cabin Use Cases: A Survey,” IEEE Intelligent Transportation Systems Magazine, February 2021.
[2] P. Nandi, A. Mishra, P. Kedia, and M. Rao, “Design of a Real-Time Autonomous In-Cabin Sensory System to Detect Passenger Anomaly,” IEEE Intelligent Vehicles Symposium (IV), Las Vegas, USA, October 20-23, 2020.
[3] S. Kusuma, J. Divya Udayan and A. Sachdeva, "Driver Distraction Detection using Deep Learning and Computer Vision," 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies, pp. 289-292, 2019.
[4] Q. Xiong, J. Lin, W. Yue, S. Liu, Y. Liu and C. Ding, "A Deep Learning Approach to Driver Distraction Detection of Using Mobile Phone," 2019 IEEE Vehicle Power and Propulsion Conference (VPPC), pp. 1-5, 2019.
[5] “Behavioral Recognition for Shared and Autonomous Mobility,“ [Online]. Available: https://www.viisights.com/wpcontent/uploads/2020/12/viisights-in-vehicle-datasheet-2021.pdf
[6] “Driver Monitoring and In-Cabin Monitoring with a Single 3D Time-of-FlightCamera,” [Online]. Availabl :www.melexis.com
[7] Ashutosh Mishra, Jinhyuk Kim, Dohyun Kim, Jaekwang Cha, Shiho Kim, “An Intelligent In-cabin Monitoring System in Fully Autonomous Vehicles,” International SoC Design Conference (ISOCC), Yeosu, Korea (South), 21-24 Oct. 2020.
[8] Netron : visualizer for neural network, deep learning, and machine learning models, [Online]. Availabl : https://github.com/lutzroeder/netron/releases/tag/v5.1.4
[9] Redmon, J., Divvala, S., Girshick, R., "You Only Look Once-Unified Real-Time Object Detection," In Proceedings of the IEEE conference on computer vision and pattern recognition, pp.779-788, 2016.
[10] Redmon, J., & Farhadi, A., ”YOLO9000: better, faster, stronger.” In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7263-7271, 2017.
[11] Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767, 2018.
[12] “github:YOLO-Fastest” [Online].Available: https://github.com/dog-qiuqiu/Yolo-Fastest
[13] “github:efficientnet-lite” [Online].Available: https://github.com/tensorflow/tpu/tree/master/models/official/efficientnet/lite
[14] “Mosaic data augmentation” [Online].Available: https://reurl.cc/gW6Qjb
[15] “github:darknet” [Online].Available: https://github.com/AlexeyAB/darknet