Research on Driver Drowsiness Detection Method based on Deep Learning

Shi Puwei
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka 808-0196, Japan

Kazuo Ishii
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka 808-0196, Japan
Email: shi.puwei537@mail.kyutech.jp, ishii@brain.kyutech.ac.jp

Abstract

Drowsiness driving will pose a serious threat to the lives of drivers and others. Determining the state of the driver through face recognition has the advantages of low cost and convenience. Therefore, this study deploys the face recognition model to the mobile phone, and finally realizes the recognition of the driver's Drowsiness by the front camera of the mobile phone. The whole research is divided into three parts. The first part is to train the face 68-keypoints recognition model based on the YOLO face algorithm. The second part is to deploy the trained model to the mobile phone using ONNX and NCNN. The third part is to calculate EAR and MAR using several facial key points, and finally complete the recognition of the driver's drowsiness state using EAR and MAR.

Keywords: YOLO face, ONNX, NCNN, EAR (Eye Aspect Ratio), MAR (Mouth Aspect Ratio)

1. Introduction

Nowadays, the car has become an indispensable means of transportation, but at the same time, driver drowsiness driving this dangerous driving behavior is common, the driver himself and others have brought indelible damage. Drowsiness driving is due to the driver's high-intensity, long time driving, resulting in excessive energy and physical exertion, and then the reaction ability and control level decreased, affecting the normal driving operation of the situation. The probability of a driver causing a traffic accident in drowsiness increases dramatically. In Canada, more than 20% of traffic accidents involve drowsiness driving. Surveys also show that 20% of traffic accidents in the United States involve drivers who are in drowsiness [1]. Therefore, it is important to study driver behavior and state recognition methods to alert and intervene when dangerous situations occur, to ensure the safety of drivers themselves and other people and vehicles on the road.

The current objective detection methods for drowsiness driving are mainly divided into based on driver physiological parameters; based on driver operating behavior; and based on vehicle driving information. In 2007, Yamaguchi et al. extracted amylase from the driver's saliva and determined whether the driver was in a drowsiness state by analyzing the changes in its composition [2]. In 2013, Mbouna et al. extracted the driver's eye state and pupil movement state as driver fatigue features and used SVM to classify the driver's drowsiness state, achieving a good recognition effect [3]. Mott G E et al. concluded that the angle of the driver's steering wheel and the lateral position of the vehicle had the highest correlation with drowsiness driving by comparing 87 vehicle parameters related to drowsiness driving in waking and moderate fatigue states [4].

As far as current research is concerned, recognition and early warning through physiological parameters is easy to interfere with normal driving and affect the drowsiness change process; identification through vehicle driving information is subject to factors such as driver skills and road conditions, and the recognition accuracy is relatively low. The drowsiness recognition method based on the driver's behavior and state has the characteristics of low equipment cost and non-invasiveness. At present, the recognition of driver characteristics is mainly concentrated in areas such as eyes, mouth and so on. At the same time, the use of cameras for face recognition also needs to consider the impact of changes in light on the recognition ability of the model.

Methods of drowsiness recognition based on face can be divided into two types. One is based on the entire image. In this method, the model will infer and recognize the entire image, and then crop the different facial feature areas. The other is based on facial feature points. The model will annotate the specified number of key points on the entire image, and then infer facial features based on the relationship between key points. Currently, the most common key points annotations include 98 key points, 68 key points, and 29 key points. This study chooses to use the most common 68 key points for model modification and dataset building.

In summary, this study aims to recognize the facial feature points of drivers through computer vision methods, establish a driver fatigue recognition method, and complete the deployment of the model on the smartphone side.

2. YOLO Face-based Face Recognition Model

2.1. *Yolo Face*

YOLO is a classic one-stage target detection algorithm based on CNN (Convolutional Neural Network). The original YOLO algorithm does not contain key points detection. It can only annotate target objects on images with bounding boxes. Later, some authors developed YOLO pose for human pose detection based on YOLO. YOLO pose adds 17 human body key points detection based on YOLO. The YOLO face used in this study is based on YOLO pose, changing the 17 key points to five key points regression. YOLO face treats face detection as a general target detection task. Therefore, its network structure is not much different from the overall YOLO pose, using only some special network modules to replace the modules in the original structure [5].

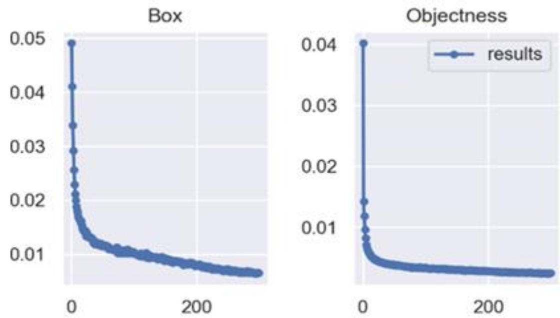
The original YOLO face includes only 5 key points detection. This study chooses to use 68 key points for extracting face-related features. Therefore, the original YOLO face needs to be modified. The modification is mainly reflected in the data structure. The data structure of each target in the original is a vector of length 16. The length of data vector modified is 138.

2.2. *Dataset for training*

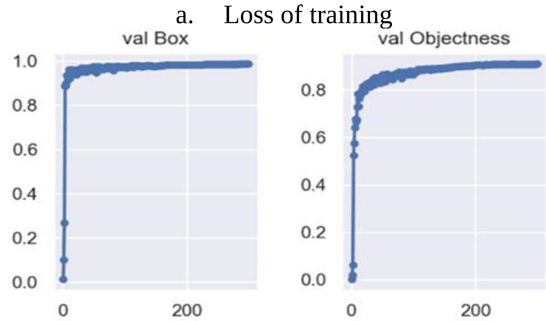
This study combines the 300w-3D dataset and the LaPa dataset to build training and validation sets for the face recognition model [6], [7]. Among them, the 300w-3D dataset has the characteristics of large data volume, accurate annotation, gender balance, age distribution balance, and bright/dark balance. However, the open/closed eye data in the 300w-3D dataset is not balanced, especially the closed eye data is relatively small. This will lead to poor model generalization ability, and eventually fail to detect closed eye data. Therefore, this study screened out the closed eye data from the LaPa dataset and combined it with the 300w-3D dataset to form a new dataset. The model is trained and validated using this new dataset.

2.3. *Training Result*

YOLO-tiny is selected as the network structure. After training for 300 epochs, the model has basically converged (Fig. 1).



a. Loss of training



b. Precision of validation

Fig. 1. Result of training

The detection performance of the model is evaluated in terms of mAP (Mean Average Precision). mAP@0.5 represents the change of mAP when the IOU threshold is 0.5, and mAP@0.5:0.95 represents the change of mAP when the IOU threshold gradually increases from 0.5 to 0.95. These two mAPs are both greater than 0.9 after the end of training, and the training effect is good (Fig. 2).

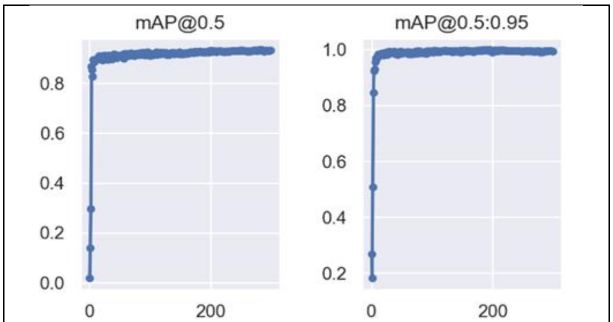


Fig. 2. mAP (Mean Average Precision)

To verify that the model still has good detection capabilities under different lighting conditions, several images are selected, and new images were obtained by decreasing the average brightness of every image by 50 in turn, and they were sent to the model. The results show that the model also has good detection ability in low brightness conditions (Fig. 3).

Fig. 3. Face Detection under Different Brightness

3. Android Smartphone Model Deployment

This study uses ONNX and NCNN as "intermediate representations" to build an inference engine based on the NCNN framework to deploy the trained face recognition model on Android phones.

In general, deep learning models rely on related frameworks and environments, which are not suitable for installation in mobile phones and other environments. In addition, deep learning models require a lot of computing power, which is difficult to run on phones without optimization. ONNX and NCNN as "intermediate representations" can solve this problem. ONNX (Open Neural Network Exchange) was jointly released by Facebook and Microsoft in 2017 for standard description of model structure and parameter calculation graphs. It can optimize network structures and reduce environmental dependencies. NCNN is a lightweight neural network inference framework developed by Tencent for mobile platforms [8], [9]. In short, the trained model needs to be converted into ONNX form and then into NCNN form before it can be deployed to a smart phone. (Fig.4)

Fig. 4. Model transformation process

The effect is shown in Fig. 5, the experimental device is an Android phone equipped with Snapdragon 888Soc (in strong light and low light environments, respectively).

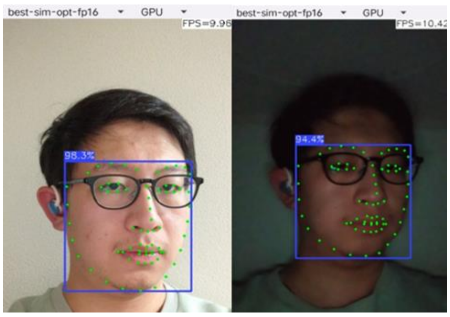


Fig. 5. The effect of the model running on a mobile phone

4. Drowsiness Detection Method by face key points

4.1. *Eye and mouth state recognition based on EAR and MAR*

After completing the extraction of facial feature regions or feature points, it is necessary to study the relationship between the open and closed state of the eyes and mouth and drowsiness based on these characteristics. The relationship was investigated in this study using EAR (Eye Aspect Ratio) and MAR (Mouth Aspect Ratio). They are based on the same principle, the value of them can reflect the opening and closing status of the eyes and mouth visually. Taking eye as an example, among the 68 facial feature points, each eye occupies 6 feature points, distributed as shown in Fig. 6. [10]

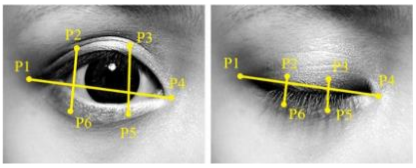


Fig. 6. Feature points of eye

When the eyes are open and closed, the relative position of P1~P6 changes greatly, especially the vertical coordinate changes between the P2/P6 and P3/P5 point pairs are more obvious, and the horizontal coordinate distance between the P1/P4 point pairs basically does not change when the eye state changes. According to the change law of the relative distance between the 6 feature points when the eye state changes, the current state of the eye can be clearly extracted, and the specific calculation is as Eq. (1).

$$EAR = \frac{\|P2 - P6\| + \|P3 - P5\|}{2\|P1 - P4\|} \qquad (1)$$

4.2. *Experiment for drowsiness detection based on EAR and MAR*

This study uses the dataset from NTHU Computer Vision Lab for related experiments and analysis. EAR and MAR are computed for each frame of all the videos in the training subset of the dataset. To minimize the effect of noise on the data, exponential smoothing filter and Gaussian filter are applied to the EAR and MAR of each sample respectively. The EAR and MAR graphs can be obtained as shown in Fig.7 and Fig. 8. The downward spike in the graph of EAR is the process of winking, and the upward spike with a certain width in the graph of MAR is the process of yawning.

By analyzing 142 video samples (71 drowsiness samples and 71 non-drowsiness samples), it can be found that the change in the frequency of winking does not reflect the drowsiness state, because there are individual differences in winking habits, and the frequency of winking rises in the drowsiness state of some people and decreases in other people. But there is always a longer period of eye closing when people are in the drowsiness state. Therefore, the presence of significantly longer than normal winking time for eye closing time can be used for drowsiness detection (Fig. 9). In addition, for MAR, the n-sigma rule can be used to distinguish between yawning and normal speaking (Fig.7 and Fig. 10). The existence of yawning can also be used for drowsiness detection. In summary, drowsiness detection methods can be constructed based on yawning and longer eye closing time.

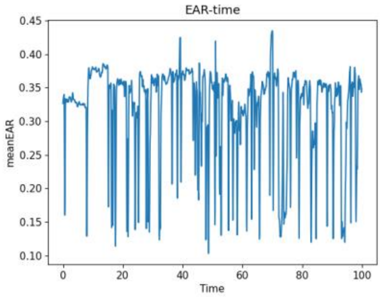


Fig. 7. EAR of a non-drowsiness sample

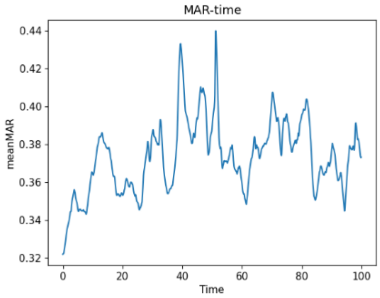


Fig. 8. MAR of a non-drowsiness sample

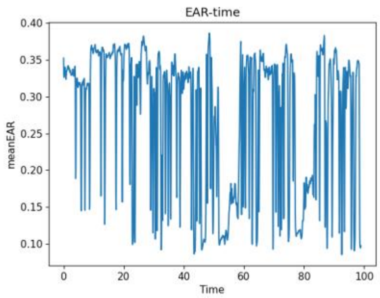


Fig. 9. EAR of a drowsiness sample

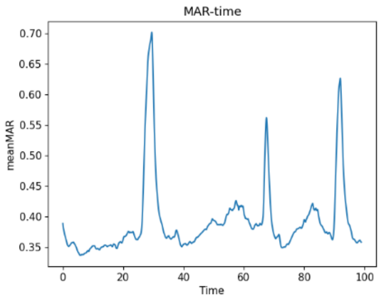


Fig. 10. MAR of a drowsiness sample

5. Conclusion

In this study, a 68-keypoints face recognition model was implemented based on YOLO face, after which the model was deployed to Android phones through ONNX and NCNN frameworks, and finally EAR and MAR were calculated by key points and their relationship with driver drowsiness driving was studied. According to the experiment, it was concluded that based on yawning and longer eye closing time can be used as a standard for drowsiness determination. In the future, it is planned to add this standard into the Android inference engine, and the network structure of the face recognition model is adjusted to improve performance.

References

1. Li Xiaoxing, Deep Learning Based Fatigue Driving Detection Methods, master's thesis, University of Science and Technology of China, 2020.
2. Yamaguchi M, Deguchi M, Wakasugi J, Hand-held monitor of sympathetic nervous system using salivary amylase activity and its validation by driver fatigue assessment, Biosensors& Bioelectronics, 2006, Vol. 21(7), pp. 1007-1014.
3. Mbouna R O, Kong S G, Chun M G, Visual Analysis of Eye State and Head Pose for Driver Alertness Monitoring, IEEE Transactions on Intelligent Transportation Systems, 2013, Vol. 14(3), pp. 1462-1469.
4. Mott G E, Forsman P, Short K R, Efficient driver drowsiness detection at moderate levels of drowsiness, Accid Anal Prev, 2013, Vol. 50(1), pp. 341-350.
5. Delong Qi, Weijun Tan, Qi Yao, YOLO5Face: Why Reinventing a Face Detector, arXiv:2105.12931(CS/CV), 2021, https://arxiv.org/abs/2105.12931.
6. Xiangyu Zhu, Zhen Lei, Xiaoming Liu, Face Alignment Across Large Poses: A 3D Solution, http://www.cbsr.ia.ac.cn/users/xiangyuzhu/projects/3DDFA/main.htm
7. Yinglu Liu, Hailin Shi, Hao She, A New Dataset and Boundary-Attention Semantic Segmentation for Face Parsing, The Thirty-Fourth AAAI Conference on Artificial Intelligence, 2020, pp. 11637-11644.
8. Open standard for machine learning interoperability, https://github.com/onnx/onnx/tree/main
9. A high-performance neural network inference framework optimized for the mobile platform, https://github.com/Tencent/ncnn
10. Tereza Soukupová, Real-Time Eye Blink Detection using Facial Landmarks, 21st Computer Vision Winter workshop, Slovenia, 2016.

Authors Introduction

Mr. Shi Puwei

He received his Bachelor's degree in Automation from Guangxi University, Institute of Electrical Engineering in China.
He is currently a master student in Kyushu Institute of Technology, Japan

Dr. Kazuo Ishii

He received his Ph.D. degrees from Tokyo University, Japan, in 1996. In 2011, he joined Kyushu Institute of Technology, where he is currently a Professor of the Department of Human Intelligence Systems.
His research interests include Underwater Robot, Neural Network and Neural Network.