Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/6/20230513

Driver fatigue detection based on lightweight MobileNetV2

Haoyuan Wang

Department of Software Engineering, Zhuhai College of Science and Technology, Zhuhai, 519000, China

15010140132@xs.hnit.edu.cn

Abstract. Fatigue driving brings a huge potential threat to the property and life safety of drivers and pedestrians. In order to improve the intelligence level of driver fatigue driving detection, a driver fatigue detection method based on MobileNetV2 lightweight separable convolutional neural network is proposed in this study. Firstly, the driver images in the data are pre-processed to make the data compatible with the model and improve accuracy, then the MobileNetV2 network model is used to detect and extract key features (mainly eye and mouth information) from the face, and finally the softmax function is used to classify the acquired features (based on the trained parameters, which have been obtained to discriminate fatigue) and determine whether the driver is in a fatigue state. The experimental results show that the proposed method has the satisfactory performance in terms of the accuracy, can basically adapt to driving scenes under different lighting, and has high robustness.

Keywords: computer vision, fatigue detection, MobileNetV2, transfer learning.

1. Introduction

As people's living conditions gradually rise and major cities' transportation networks continue to be improved, the total number of cars in China is also increasing, which brings convenience to people's travel, but also leads to frequent traffic accidents, posing a significant risk to the property and life safety of drivers and pedestrians. Statistics show that roughly 20% of all traffic accidents are caused by drivers who are fatigued, and about 40% of major casualties are caused by fatigue driving [1]. Therefore, the research of driver fatigue detection is of great significance to reduce the casualties and property losses caused by traffic accidents.

Up to now, the design and development methods of fatigue driving detection system at home and abroad can be divided into subjective detection method and objective detection method [2]. The objective detection method provides higher accuracy and better real-time performance since it is not influenced by the driver's subjective consciousness. The benefits of the objective detection method have made it the primary research area for the identification of driver fatigue. There are mainly three kinds of objective detection methods: based on motor vehicle behavior characteristics, based on driver's physiological characteristics, and based on driver's facial characteristics. Facial feature-based has become popular because of its advantages of non-contact, accuracy and low cost. Zhu Feng et al. proposed to perform face detection through the combination of the improved Yolov3 algorithm and Kalman filter algorithm [3]. Based on the longest continuous period of eye closure, the frequency of yawns, and the percentage of eyelid closure over pupil over time (PERCLOS) per unit of time, a

DOI: 10.54254/2755-2721/6/20230513

boosted tree-based algorithm is used to identify facial important features. Multi-feature fusion fatigue detection uses these three features. Pan Zhigeng et al. used Adaboost algorithm for face detection [4], and divided the roughly human eye region into segments based on the arrangement of the face's "three courtyards and five eyes." A fuzzy comprehensive evaluation algorithm is used to analyze the three influencing factors of eye rectangle area aspect ratio, fitting ellipse area, and pupil melanin proportion, and to distinguish the opening and closing state of eyes. In the process of eye location, OSTU threshold segmentation, nonlinear point operation, and integral projection are used to remove eyebrows. Finally, the PERCLOS concept is used to determine the driver's level of weariness. AlexNet and the transfer learning theory were proposed to be incorporated into the research on the recognition of driver behavior States by Rong Hui [5] and others, and experiments were used to confirm the efficacy of the proposed driver state recognition algorithm for the recognition of seven different driver States.

To sum up, most current studies mainly detect fatigue by extracting overall behavior features and partial face features [6,7]. However, such a way of overall behavior or partial characteristics may have some problems, such as the loss of some key characteristics and the single fatigue index that can be referred to, so it is difficult to show the fatigue state. Therefore, based on the driver's facial features, a fatigue state recognition network is proposed, which combines eye and mouth features and Convolutional Neural Network (CNN). The method combines the two feature points of eyes and mouth to judge fatigue, and improves the fatigue detection performance through the study of ImageNet transfer learning techniques. This study effectively solve the problem of misjudgment caused by single fatigue index and missing features. This paper aims to provide a scientific reference for the intelligent development of driver fatigue detection through the fatigue recognition method based on multi-feature of driver face fusion.

2. Method

2.1. Dataset description and preprocessing

The data set used is the data set Drowsiness _ dataset from kaggle [8]. The data collection contains 2,900 images. Additionally, it is separated into four groups: Closed, Open, Yawn, and No Yawn. Where the yawn and no yawn images are 640x480 human face JPG images with or without yawning, and the Closed and Open images are 300x300 human eyes opening and shutting JPG images. Open and no yawn are classified as not being weary in this project's binary classification approach, whereas closed and yawn are classified as being fatigued. Additionally, 100 images from each class were chosen at random to serve as the test set. The sample photos from the compiled dataset are shown in Figure 1.

(a) yawn (b) no_yawn

Figure 1. Sample images in the collected dataset.

Provide the sample images in terms of data preprocessing operation, this paper modifies the size of the data to be processed to 56x56 for model training and testing and normalizes the pixel value from the original pixel space (0-255) to (0,1), its advantages are possible to prevent the properties of large

DOI: 10.54254/2755-2721/6/20230513

value ranges from over-dominating the properties of small value ranges. Another advantage is that that numerical complexity in the calculation process can be avoided to make the data comparable. Then the data is encoded by one-hot, and the value of the discrete feature is extended to the Euclidean space, and a value of the discrete feature corresponds to a point in the Euclidean space to make the calculation of the distance between features more reasonable.

2.2. Experimental network model and procedure

In this project, the MobileNetV2 model of CNN is applied [9]. The basic CNN consists of three structures: convolution, activation, and pooling [10]. The result of the CNN output is a specific feature space for each image. When dealing with the task of image classification, this study will take the feature space output by CNN as the input of fully connected layer or Fully Connected Neural Network (FCN), and use the fully connected layer to complete the mapping from the input image to the label set, that is, classification. Of course, the most important work in the whole process is how to adjust the network weight iteratively through the training data, that is, the backward propagation algorithm. At present, the mainstream Convolutional Neural Networks (CNNs), such as VGG and MobileNet, are adjusted and combined by simple CNN. The MobileNet V2 model, proposed by the Google team in 2018, is more accurate and smaller than the traditional CNN model. The highlights are Inverted Residuals and Linear Bottlenecks (the last layer of the structure is a linear layer). The network structure of the model is shown in Table 1.

Table 1. The network structure of MobileNetV2.

Input	Operator	t	c	n	s
$224^2 \times 3$	conv2d	——	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d 1×1	——	1280	1	1
$7^2 \times 1280$	avgpool 7×7	——	——	1	——
$1 \times 1 \times 1280$	conv2d 1×1	——	k	——	——

In Table 1, input is the size before the input feature map, bottle neck is the inverse residual network structure, t is the expansion factor, c is the depth of the output feature matrix, n is the number of bottle neck repetitions, and s is the step size.

The network consists of input layer, convolution layer, pooling layer, fully connected layer and output layer. Each convolution layer has a convolution kernel of size 3x3. The network first receives images from Drowsiness _ dataset data sets, each image is preprocessed, and then the input images are transmitted through the network. After each convolution layer, there is a 2x2 maximum pooling layer to reduce the dimension and compress the features to speed up the calculation and prevent overfitting. In order to make the sparse model can better mine the relevant features and fit the training data, the Relu activation function is used in the fully connected layer. In order to improve efficiency and accuracy, ImgaeNet pre-training set combined with MobileNet V2 model is used for training. The softmax function is then used for output classification to classify the image into fatigue and non-fatigue. The softmax function compresses (maps) the K-dimensional real vector into another

DOI: 10.54254/2755-2721/6/20230513

K-dimensional real vector, where each element in the vector has a value between (0, 1). It is often used for multi-classification problems. The formula is as follows:

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \tag{1}$$

Finally, the test set is put into the trained model for comparison, and the image classification in the test set is judged to obtain the accuracy.

2.3. Implementation details

The optimizer of choice is the Adam optimizer, because it combines the advantages of the gradient descent algorithm with adaptive learning rate and the momentum gradient descent algorithm, which can not only adapt to sparse gradients, but also alleviate the problem of gradient oscillation. In order to make the network gradient more accurate, the model is trained with the default learning rate of 0.0001 and batch _ size = 32. The number of training sessions, that is, epochs, is set to 50.

Loss is usually used to show the difference between the prediction and the data. The smaller the loss is, the better the robustness of the model is. The experiment uses categorical _ crossentropy classification cross entropy loss function, and its formula is as follows:

$$Loss = -\sum_{i=1}^{output\ size} y_i \cdot \log \hat{y}_i \tag{2}$$

According to the formula, it can be found that because y_i is either 0 or 1. When y_i is equal to 0, the result is 0, and if and only if y_i is equal to 1, there is a result. That is to say, the categorical _ crossentropy only focuses on one result, so it generally cooperates with softmax to do single-label classification.

3. Results and discussion

3.1. Results

Table 2. The performance of various models.

Method	Performance			
	training loss	training accuracy	testing loss	testing accuracy
MobileNet (No pretrained)	0.0082	0.9975	1.0094	0.7375
MobileNet（ImageNet _Weight	0.0107	0.9970	0.2533	0.9275
MobileNetV2 (No pretrained)	0.0285	0.9940	1.2509	0.7575
MobileNetV2（ImageNet _Weight	0.0052	0.9990	0.6011	0.9549

This experiment uses MobileNet and MobileNetV2 for comparison. At the same time, ImageNet pre-training weights were used to observe the experimental results (a total of four groups). In this experiment, epoch was set to 50. Finally, it can be found that the MobileNet and MobileNetV2 models with pre-training weights have very high accuracy in both the training set and the test set. The test set

DOI: 10.54254/2755-2721/6/20230513

for MobileNetV2 is the most accurate. Although the training set results of the model without pre-training weights are also very good, the accuracy of the test set is only more than 70%. Figure 2 presents the accuracy value of the pre-trained MobileNetV2 model, and Figure 3 shows the loss value during the training process.

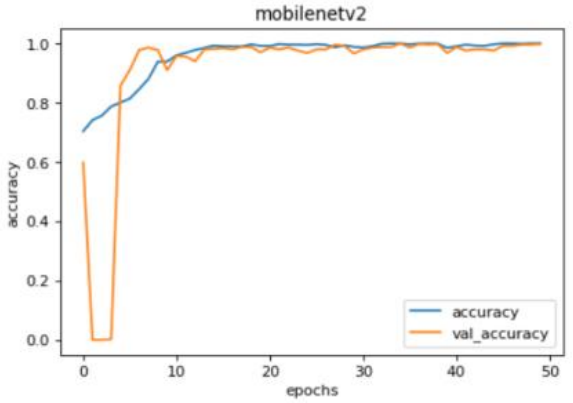

Figure 2. The accuracy of models.

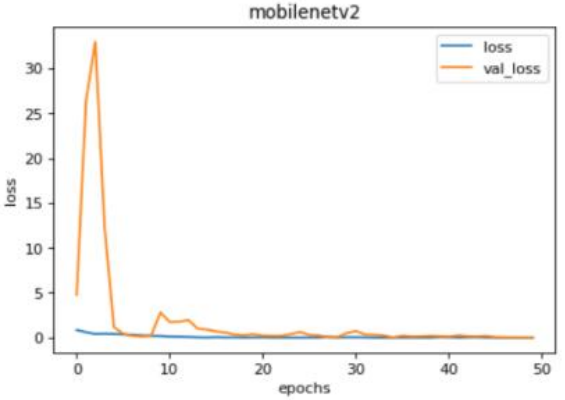

Figure 3. The loss of models.

3.2. Discussion

As can be seen in Table 2, pre-training the model will work better, and all that has to be learned to train a neural network is actually a suitable parameter. That direct training approach manifests itself in the model parameters, which is to randomly initialise all the model parameters, and then the model starts from 0 to learn the appropriate parameters for this current task.

Using a pre-trained model means that it is no longer necessary to train all parameters from 0. However, some parameters may not be appropriate for the task at hand, so it is only necessary to modify the current parameters slightly to get better results, which will not only reduce the learning time but also improve the accuracy. This experiment also shows that there are already feature parameters in the ImageNet library that are useful for learning face information and can be applied to face fatigue detection.

From the experimental results, it can be known that the accuracy of the MobileNetV2 model will be higher than that of the MobileNet model. The main difference between the two is that MobileNetV2 adds a layer of Pointwise convolution before Depthwise convolution. The number of Kernels of the Depthwise convolution in the depthwise separable convolution of MobileNet V1

DOI: 10.54254/2755-2721/6/20230513

depends on the Depth of the previous layer, that is, the Depthwise convolution itself does not have the ability to change the number of channels, and Depth is the channel. When a layer of Pointwise convolution is added, it can be up-dimensioned or down-dimensioned, as in MobileNetV2. After the Pointwise convolution is up-dimensioned, the Depthwise convolution can work in higher dimensions. The straightforward explanation is that linear convolution is employed to extract features from the low latitude space with relu output in order to minimize feature loss. To prevent excessive feature loss for layers with relu, the number of feature channels is first raised. The main facial characteristics are preserved as much as feasible in this experiment, producing a more precise test set.

4. Conclusion

The proposed fatigue detection method combines traditional features and convolutional neural networks using the MobileNetV2 model to determine driver fatigue. The proposed method not only allows for in-depth mining of overall facial features, but also takes into account local features of the face that contain rich information. The method can be adapted to driving scenarios with different lighting conditions and can guarantee a high detection accuracy. In this experiment, transfer learning is also used, using some feature parameters that may already exist in the ImageNet library to facilitate the learning of face information, and applying them to face fatigue detection in this experiment. The experimental results show that the use of the MobileNetV2 model and the addition of the migration learning technique led to better results in terms of effectiveness and robustness. As the masked driver images are not included in the dataset, further research on masked fatigue detection will be carried out in future work.

References

[1] Nanjing Report. The harm of the fatigue driving [R]. https://m.gmw.cn/baijia/2022-04/21/1302910346.html
[2] Zhang R et al. 2022 A review of research on driver fatigue driving detection methods [J]. Computer Engineering and Applications:1-19
[3] Zhu F et al. 2022 Driver fatigue detection based on improved Yolov3 [J] Science Technology and Engineering 22(08):3358-3364
[4] Liao J et al. 2016 Research on driver fatigue driving detection technology based on eye-movement features [D] Hunan University of Science and Technology
[5] Rong H et al. 2019 A method for driver behavior state recognition based on migration learning and AlexNet [J] Science Technology and Engineering 19(28):208-216
[6] Isaza C et al. 2019 Dynamic setpoint model for driver alert state using digital image processing [J] MultimediaTools and Applications 78 (14): 19543-19563
[7] You F et al. 2019 A real-time driving drowsiness detection algorithm with individual differences consideration [J]. IEEE Access.
[8] Kaggle 2020 https://www.kaggle.com/datasets/dheerajperumandla/drowsiness-dataset.
[9] Zhao Y et al. 2021 A face recognition system based on MobileNetV2 and Raspberry Pi [J] Computer System Applications 2021 30(8):67-72.
[10] Yu Q et al. 2020 Improved denoising autoencoder for maritime image denoising and semantic segmentation of USV [J] China Communications, 17(3), 46-57.