Available online www.jsaer.com

Journal of Scientific and Engineering Research, 2022, 9(4):17-26

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Distracted Driving Behavior of Operation Recognition Method Based on YOLOv5 and BPNN

Sun Longxiang, Feng Huanchao, Zhang Min, Li Anmengdie, Zhang Jinglei*

*School of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo, Shandong, China. jinglei@sdut.edu.cn

Abstract In order to accurately identify the driver's distracted driving behavior, reduce the injury of traffic accidents to the personnel in the vehicle and improve the driving safety level, this paper takes the driver's distracted driving behavior as the research object, build a distracted driving experimental environment based on real vehicles, and collected 27902 distracted driving image data of 20 drivers, The combined recognition model of YOLOv5 and BPNN was constructed, the video frame image was input into the trained YOLOv5 model, the boundary box data of the driver's left, right hand and face were output, and then the boundary box data was input into the BPNN model for distracted driving behavior recognition. The precision, recall and F1 score of YOLOv5 and BP NN combined recognition model for operation distraction are 0.926, 0.930 and 0.928 respectively, and the overall macro F1 of the model is 0.938. Compared with other similar studies, this model has stronger recognition performance. The research on distracted driving recognition method in this paper provides a certain theoretical basis and method for the improvement of vehicle active safety early warning and safety assisted driving system, and is of great significance to improve the level of road safety.

Keywords Distracted driving; Object detection; Classification model; Driving behavior recognition

1. Introduction

Among the factors affecting the level of driving safety, distracted driving has become an important factor that causes traffic accidents. The United States Highway Safety Administration (NHTSA) defines distracted driving as any activity that distracts driving attention, and divides distracted driving into visual distraction, operational distraction and cognitive distraction [1]. In 2018, there were 2628 fatal crashes involving distraction on U.S. roads, resulting in 2841 deaths (8% of the total deaths in traffic accidents), while this figure rose to 3142 in 2019, an increase of 10.6% [2]. Therefore, we urgently need to find an effective measure to reduce traffic accidents caused by distracted driving, which is very important to improve the level of driving safety [3].

In recent years, a large number of experts and scholars have studied distracted driving from different directions, mainly including driver visual characteristics, physiological signal characteristics, vehicle operation information and image recognition and detection. Because eye tracker can effectively obtain human visual characteristic data, it is used by many scholars to study distracted driving from the visual direction. Sodhi et al. used eye tracker to record the relevant visual parameters when the driver adjusted the radio, observed the rearview mirror and other sub task operations, and then analyzed the influence law of sub task operation on driver distraction [4]. Omid dehzangi proposed a wearable physiological sensor to quantify skin conductivity (SC), to describe and identify distractions in natural driving. By using embedded random forest feature selection and set bag classifier, using 10-d and 15-d feature space, the improved accuracy is 92.9% and 93.5% respectively [5].

Osama A. Osman proposed a two-tier hierarchical classification method using machine learning to identify different types of secondary tasks performed by drivers using their driving behavior parameters [6].

In addition, benefiting from the development of object detection and deep learning technology, it is easier to identify the distracted state of the driver by obtaining video image data through the camera, which could avoid the interference of the sensor in direct contact with the driver. Duy tran implemented and evaluated four deep convolution neural networks on the embedded graphics processing unit platform, including VGG-16, AlexNet, GoogleNet and ResNet. The research result shows that GoogleNet is the best of the four distraction detection models on the test platform of driving simulator [7]. Xuli Rao proposed a distracted driving recognition method based on deep convolution neural network [8]. This method uses principal component analysis technology to whiten the driving image and reduce the redundancy and correlation of pixel matrix. A multi-layer CNN network is constructed and its key parameters are optimized. The recognition accuracy of the model is 97.31%. Li Li, Boxuan Zhong and others used YOLOv3 model to detect the driver's right ear and right hand at the same time, and input the detected image into a multi-layer perceptron. For the overall distraction detection, F1 value reached 0.74 [9].

In this paper, through the real vehicle experiment of distracted driving, five different distracted operations are collected, and built the distracted driving image dataset, presents a distracted driving behavior recognition method based onYOLOv5 and BPNN This method achieves good detection results on the self-built datasets.

2. Experiment and Data

2.1. Equipment and Environment

In this experiment, the distracted driving image data of drivers are collected based on the real vehicle and real road environment. Considering the certain risk of distracted driving, this paper selects a closed road for the experiment to avoid the danger to other vehicles and pedestrians in the process of the experiment. The experimental road is a two-way two lane road with a length of about 2km, including straight-line driving section and right angle turning section. The equipment involved in the whole experimental process of this paper includes Haval H7 SUV and Logitech c310-720p external camera.

The camera is erected at the lower end of the right A-pillar in the vehicle cab, as shown in position ‘A’ in Figure 1. The camera faces the driver and can shoot the actions of the driver's upper body and upper limbs.

2.2. Experimental contents

20 drivers were recruited in this experiment, including 15 men and 5 women, all holding C1 motor vehicle driver's license. The average age of drivers was 25.3 years and the average driving age of drivers was 2.3 years. The duration of the single experiment was 3 minutes. The distracted driving experimental control group and experimental group were set up respectively. The control group was the driver who drove normally and kept driving, and the experimental group was the driver who operated distracted driving. The control group lasted 2 minutes and the experimental group lasted 1 minute. The driver performed the distraction task according to the voice prompt. The experimental process is as follows:

1) After entering the vehicle, the driver is familiar with the environment in the vehicle and tells the driver of the five operation distraction tasks that need to be made in the next experiment. The specific tasks are shown in Table 1.

Table 1: Distracted driving task

Number	**Operational distraction**
1	Eat or drink
2	Operate the radio in the center console
3	Make a phone call
4	Hold mobile phone to view information or send a text
5	Turn around and pick up the items in the back

2) The driver drives the vehicle according to the voice prompt, drives the vehicle according to the established route, maintains the speed below 50km / h, and carries out a control experiment for 2 minutes to collect driver behavior data under normal driving conditions.
3) The driver makes the corresponding operation distraction task according to the co pilot's prompt. Considering the experimental time and safety problems, each group of experiments selects only 3 of the 5 operation distraction tasks for operation. The driver drives normally every time he completes an operation task and waits for the co pilot to prompt the next task.
4) After completing the distraction task, the driver shall slow down after hearing the voice prompt of "end of experiment", find a safe place to stop, and the single experiment ends.

2.3. Data presentation

A total of 20 videos were obtained in the experiment, and then the video frames were intercepted at an interval of 0.2S. After eliminating the invalid images such as overexposure and darkness, a total of 18754 images were intercepted, of which 13962 images were normal driving and 4792 images were distracted driving. An example of intercepted driver behavior is shown in Figure 2.

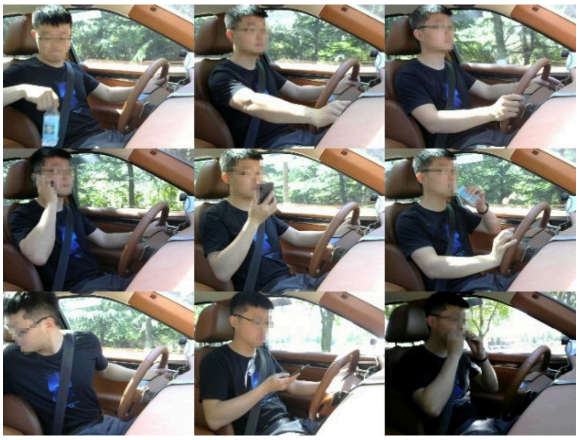

Figure 2: Examples of driver's distracted behavior

3. Construction of Model

The distracted driving recognition model in this paper includes two parts. The first part is to input the intercepted image into YOLOv5 model, detect the driver's left hand, right hand and face position, and output the detection results as a file of detection frame position data. The second part input the boundary box position data into the BPNN model and output the distracted driving types of drivers.

3.1 YOLOv5

YOLOv5 is a powerful target detection algorithm based on pytorch framework. Its network structure is divided into four parts: Input, Backbone, Neck and Output. The network structure is shown in Figure 3.

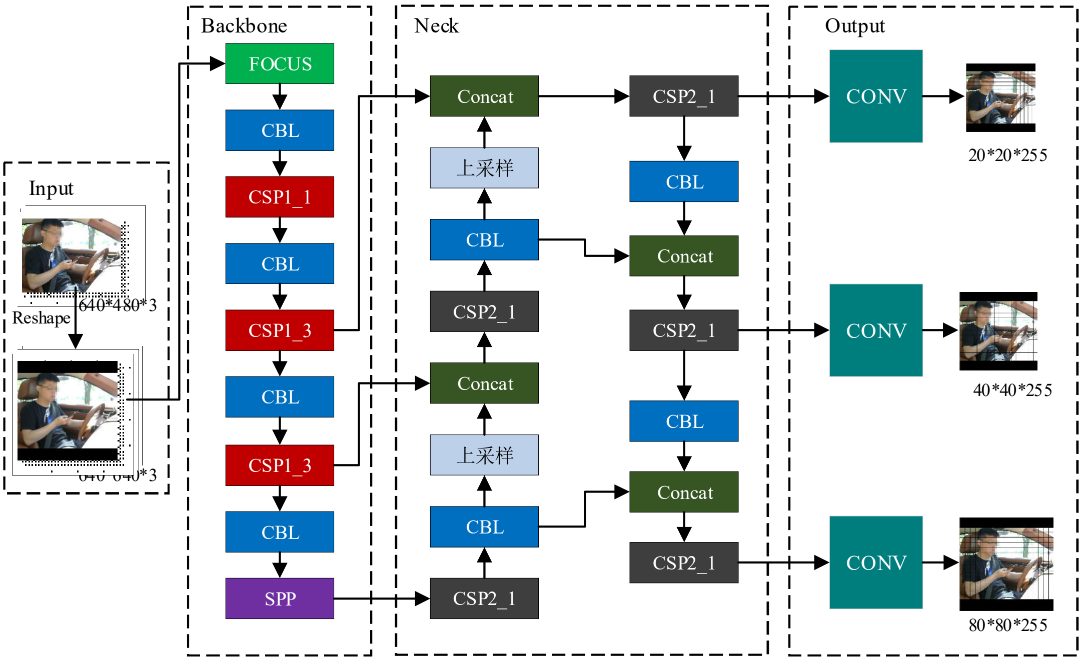

Figure 3: YOLOv5 network structure

The input of YOLOv5 adopts mosaic data enhancement technology, random scaling, random cutting and random arrangement, which greatly enriches the detection data set and makes the network more robust; The use of adaptive image scaling can effectively reduce the amount of reasoning calculation and improve the detection speed.

YOLOv5 is composed of Focus structure and CSPNet; In the Focus structure, After 4 Slice operations and 1 convolution operation with 32 convolution kernels, the image is transformed from the original 640×640×3 becomes 320×320×32[10]. In order to reduce the amount of calculation and ensure the accuracy, CSPNet integrated the change of gradient into the characteristic diagram from beginning to end by using the idea of dense cross layer hop connection [11].

In order to strengthen the ability of network feature fusion, the CSP2 structure is designed based on CSPNet in the Neck part. Finally, in the output layer, the loss function IoU_Loss of yolov3 is replaced by GIoU_Loss as the loss function of the boundary box, which solves the situation that the gradient cannot be optimized when there is no overlapping target, and can reflect the way in which the two targets overlap [12]. The difference between IoU and GIoU is shown in formula (1), formula (2) and Figure 4.

$$IoU_{A,B} = \frac{S_A \cap S_B}{S_A \cup S_B} \tag{1}$$

$$GIoU_{A,B} = IoU_{A,B} - \frac{S_C - (S_A \cup S_B)}{S_C} \tag{2}$$

S_A is the area of target frame a, and S_B is the same;

S_C is the area of the smallest box C that can cover a and B.

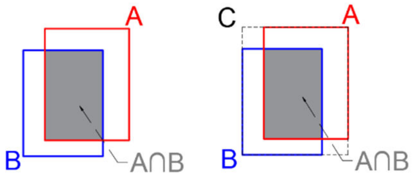

Figure 4 Differences between IoU and GIoU

3.2 BPNN

BPNN (back propagation neural networks) is a widely used multilayer neural network. It is composed of input layer, hidden layer and output layer. The hidden layer can be one or multiple. Each layer of the network is composed of multiple neurons. The neurons between the front and back layers are fully connected, and the neurons in the same layer are not connected. BPNN algorithm is a supervised learning algorithm. In BPNN, after the data enters the neuron, its activation value will propagate from the input layer through each hidden layer to the output layer, and then back propagate according to the principle of reducing the error between the expected value and the output value, return from the output layer through the hidden layer to the input layer, correct the weight layer by layer, and finally control the error within the required range after multiple training to obtain high accuracy.

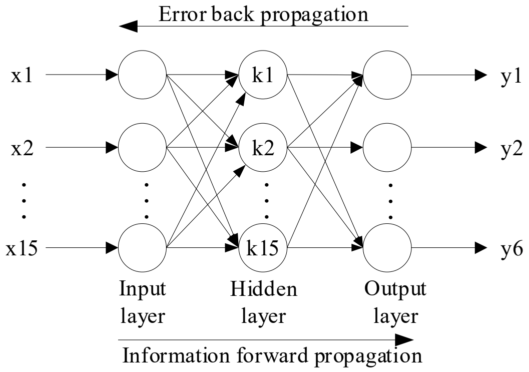

Figure 5: Schematic diagram of BPNN network model

This paper constructs a three-layer BPNN network model, with one input layer, one hidden layer and one output layer. The input layer contains 15 nodes, corresponding to 15 dimensional features such as the position data and label number of the detection frame of the driver's left hand, right hand and face; The hidden layer contains 15 nodes, and the number of hidden layer nodes is determined according to formula (3); The output layer contains 6 nodes, which correspond to 5 different operations, distracted driving behavior and normal driving. The network structure is shown in Figure 5.

$$M = \sqrt{m+n} + \alpha \tag{3}$$

m : Number of neurons in input layer;

n : Number of neurons in the output layer;

α : Is a constant between [0, 10]. In this paper, when calculating the number of neurons in the hidden layer, α =10.

In this paper, the forward propagation of BPNN network adopts the Linear function as the transfer function. The Linear function is shown in formula (4).

$$f(net) = k \cdot net + c \tag{4}$$

In formula (4), net is the network input and c is the constant.

The back propagation adopts the ReLU function as the transfer function, and the ReLU function is shown in formula (5).

$$Relu(x) = \begin{cases} x \ , & x > 0 \\ 0 \ , & x \le 0 \end{cases} \tag{5}$$

This study classifies various distracted driving behaviors, so the multi classification cross entropy loss function is adopted. The function expression is shown in formula (6).

$$L = \frac{1}{N} \sum_{i} \sum_{c=1}^{M} y_{ic} \log\left(p_{ic}\right) \tag{6}$$

M is the number of categories;

y_{ic} is a symbolic function (0 or 1), if the real category of the sample is equal to c, take 1, otherwise take 0;

p_{ic} is the prediction probability that the observation sample belongs to category c.

The optimizer selects Rprop, the learning rate in the training process is 0.02, and the number of iterations is 200.

4. Results

4.1 Results of YOLOv5

The training and testing of the model are based on pytorch framework, and the hardware environment for training and testing is windows10 professional 64-bit platform. The computer processor model is Intel Core i5-9600K, and the processor frequency is 3.70GHz. The memory capacity of the computer is 16GB and the flash memory frequency is 2666mhz. In order to speed up the training and testing, the model training in this paper uses GPU to accelerate the calculation. The model of the graphics card is NVIDIA Titan XP. The graphics card has 3840 CUDA cores (stream processor) and 12GB memory capacity, which can well support GPU accelerated computing.

In this paper, 1500 labeled images are trained. The ratio of training data to test data is 8:2, that is, 1200 images are used for training and 300 images are used for testing the model. The number of training is 40 epochs, and the resolution of the input image is 640. The training results are shown in Table 2 and Figure 6.

Table 2: Training results of YOLOv5 model

Parameter	Left hand	Right hand	Face
Precision	0.993	0.987	0.999
mAP 0.5:0.95	0.865	0.819	0.978
mAP 0.5	0.992		
Speed	17fps		
Model size	170.98MB		
Training duration	1.67hours		

mAP: mean Average Precision.

mAP0.5:0.95: Average over different IoU thresholds.

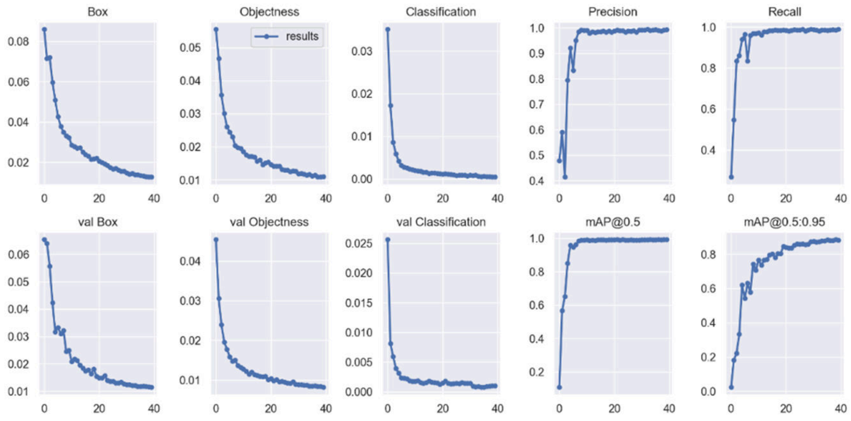

Figure 6: YOLO v5x model training result curve

Box: is the mean value of GIoU loss function;

Objectness: is the mean loss of target detection;

Classification: is the mean value of classification loss;

Precision: The proportion of the part that the classifier considers to be a positive class and is indeed a positive class in all classifiers. The calculation formula is shown in formula (7).

Recall: The proportion of the part that the classifier considers to be positive and indeed positive in all positive classes. The calculation formula is shown in formula (8).

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

TP: true positive;
FP: false positive;
FN: false negative.

It can be seen from Figure 6 that in the train set, the mean value of GIoU loss function (box) gradually decreases with the increase of epoch, and the mean value of object detection loss (objectness) and classification loss (classification) also gradually decrease with the increase of epoch, and finally maintain a relatively stable low value in the fluctuation. Precision and recall gradually increase with the increase of epoch, and finally maintain a relatively stable high value in the fluctuation. In the verification set, the changes of GIoU loss function mean (box), target detection loss mean (objectness) and classification loss mean (classification) are roughly the same as those in the training set. map0.5 and map0.5: 0.95 gradually approaches 100% with the increase of epoch, and finally reaches 99.2% and 88.7% respectively.

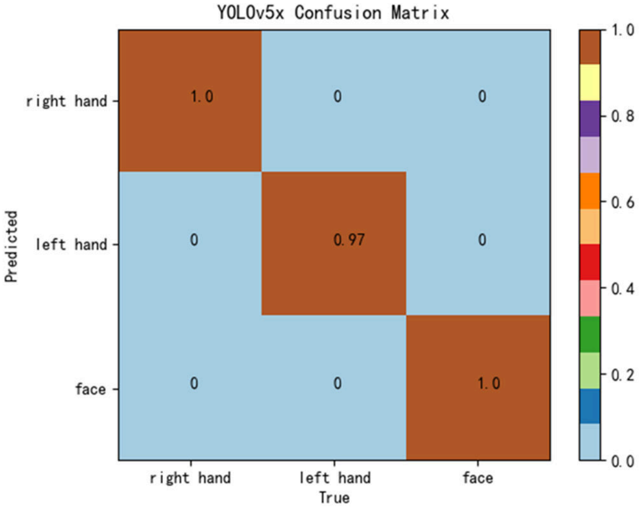

Figure 7: YOLO v5 detection confusion matrix

It can be seen from Figure 7 that the detection precision of YOLOv5 model for the driver's right hand, left hand and face reaches 100%, 97% and 100% respectively. It can be seen that the detection precision is very high. The schematic diagram of partial test image detection is shown in Figure 8.

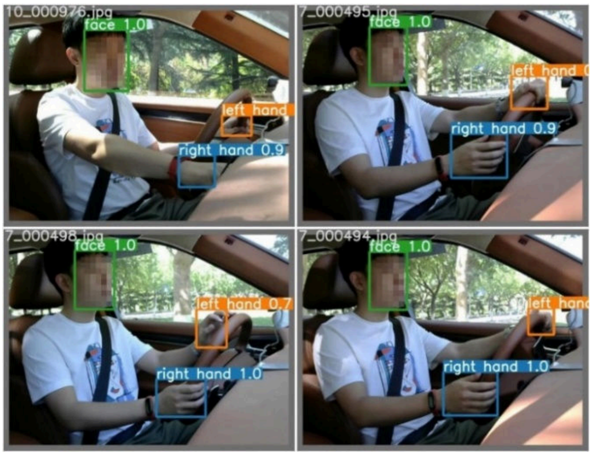

Figure 8: Schematic diagram of partial test image detection

4.2 Results of BPNN

The dataset is divided into 18754 groups of 15 dimensional data including the driver's right hand, left hand and face positions detected by YOLOv5 model. 80% of the data is used for the training of BPNN network model, called the training set, and 20% of the data is used to test the trained network model, called the test set. In the test set, there were 3751 groups of data, including 2788 groups of normal driving data, 360 groups of Hold mobile phone to view information or send a text, 266 groups of making a phone call, 137 groups of operating the central console radio, 102 groups of drinking or eating, and 98 groups of turn around and pick up the items in the back.

In the process of model training, the loss value is recorded every 5 iterations, and the final loss value reaches 0.0988. The change curve of training loss value is shown in Figure 9. After inputting the test set data into the trained BPNN network model, the classification report shown in Table 3. The training time of the model is 83 seconds and the test time is 16 seconds.

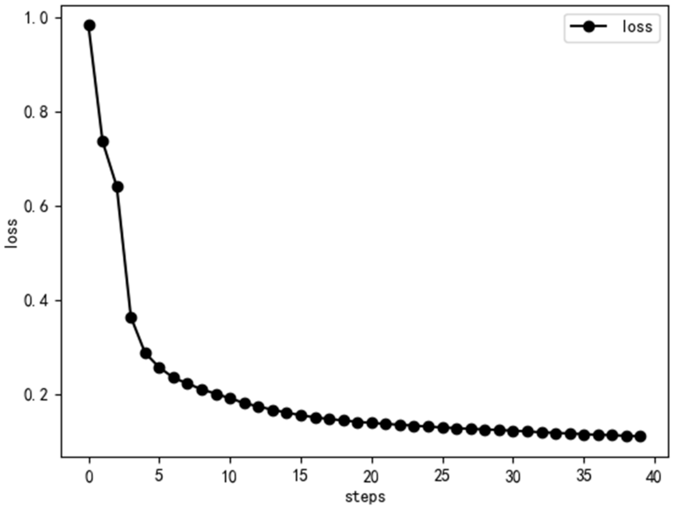

Figure 9: BPNN training loss value

Table 3: Operation distraction classification report of BPNN model

Label	Operational distraction	Precision	Recall	F1-score	Support
0	Drive normally	0.99	0.99	0.99	2788
1	Eat or drink	0.95	0.94	0.95	360
2	Operate the radio in the center console	0.86	0.94	0.90	266
3	Make a phone call	0.94	0.97	0.95	137
4	Hold mobile phone to view information or send a text	0.94	0.93	0.94	102
5	Turn around and pick up the items in the back	0.94	0.87	0.90	98
Average of operational distraction categories		0.926	0.930	0.928	963
Average of all categories		0.937	0.940	0.938	3751

Finally, 3751 groups of data obtained an overall classification accuracy of 0.97. The calculation formula of accuracy is shown in formula (9). The average precision, recall and F1 score of all categories are 0.937, 0.940 and 0.938 respectively. From table 3, we can find that except for the distracted operation such as making a phone call corresponding to label 2, the prediction precision of other types of distracted operation exceeds 0.94. Combined with the classification confusion matrix in Figure 10, 23 groups of normal driving data are predicted to make a phone call, and 6 groups of Turn around and pick up the items in the back are predicted to make a phone call, The precision of the opponent's phone recognition is 251 / (23 + 5 + 251 + 33 + 4 + 6) = 0.86, which is lower than other categories.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

TP: true positive;
TN: true negative;
FP: false positive;
FN: false negative.

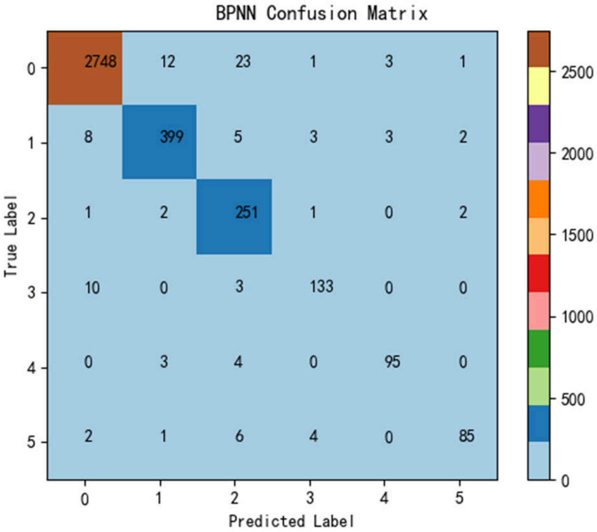

Figure 10: Distracted driving classification confusion matrix

Although the overall classification accuracy of the model is high, the analysis shows that the proportion of normal driving categories of the model is much higher than that of other categories, and the precision, recall and F1 score of the model for normal driving prediction are also very high, which leads to the distortion of the overall classification accuracy of the model to a certain extent, Therefore, in order to more accurately reflect the recognition performance of the model on the category of distracted driving, this paper calculates the average value of the evaluation index of distracted driving corresponding to labels 1-5. The average precision, recall and F1 score are 0.926, 0.930 and 0.928 respectively, which are lower than the average value of the evaluation index of all categories in Table 3.

Due to the unbalanced proportion of all kinds of distraction data in the operation distraction data set and the large difference in the number, especially the large number of normal driving categories, the performance of the model cannot be evaluated only based on accuracy. F1 score integrates the output of precision and recall, which can reflect the model performance of classification unbalanced data sets at the same time, the BPNN model established in this paper is a multi-classification model, macro F1 needs to be calculated to evaluate the performance of the model.

F1 score, also known as balanced f score, is defined as the harmonic average of precision and recall. The calculation formula of F1 score is shown in formula (10). Macro F1 is the average value of F1 score in all classes. The calculation formula is shown in formula (11).

$$F1-score = 2\times\frac{Precision\times Recall}{Precision+Recall} \tag{10}$$

$$Macro\quad F1 = \frac{\sum_{i=0}^{N-1} F1-score_i}{N} \tag{11}$$

N is the number of categories.

According to equation (11), macro F1 of BPNN operation distracted driving recognition model is (0.99 + 0.95 + 0.90 + 0.95 + 0.94 + 0.90) / 6 = 0.938.

5. Conclusion

In this paper, various evaluation indexes such as accuracy, precision, recall, F1 score and macro F1 are introduced to evaluate the model. From the evaluation indexes, the driver operation distraction recognition model based on YOLOv5 and BPNN has strong classification performance, can accurately identify a variety of different operation distraction types, and the multi classification recognition performance of the model is good, it provides a reliable method for the identification of drivers' distracted driving behavior, which can reduce the traffic safety hazards caused by distracted driving to a certain extent.

References

[1]. National Highway Traffic Safety Administration. Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices [J]. Washington, DC: National Highway Traffic Safety Administration (NHTSA), Department of Transportation (DOT), 2012.

[2]. National Highway Traffic Safety Administration. Distracted Driving 2018 [J]. Report No. DOT HS 812 926.

[3]. National Highway Traffic Safety Administration. Distracted Driving 2019 [J]. Report No. DOT HS 813.111.

[4]. M. Sodhi, B. Reimer, J. L. Cohen, E. Vastenburg, R. Kaars, S. Kirschenbaum. On-road driver eye movement tracking using head-mounted devices[P]. Eye tracking research & applications,2002.

[5]. Omid Dehzangi, Vaishali Sahu, Vikas Rajendra, Mojtaba Taherisadr. GSR-based distracted driving identification using discrete & continuous decomposition and wavelet packet transform[J]. Smart Health, 2019,14:

[6]. Osama A. Osman, Mustafa Hajij, Sogand Karbalaieali, Sherif Ishak. A hierarchical machine learning classification approach for secondary task identification from observed driving behavior data[J]. Accident Analysis and Prevention,2019,123:

[7]. Duy Tran, Ha Manh Do, Weihua Sheng, He Bai, Girish Chowdhary. Real-time detection of distracted driving based on deep learning[J]. IET Intelligent Transport Systems,2018,12(10):

[8]. Rao Xuli, Lin Feng, Chen Zhide, Zhao Jiaxu. Distracted driving recognition method based on deep convolutional neural network[J]. Journal of Ambient Intelligence and Humanized Computing,2019,12(1):

[9]. Li Li, Boxuan Zhong, Clayton Hutmacher, Yulan Liang, William J. Horrey, Xu Xu. Detection of driver manual distraction via image-based hand and ear recognition[J]. Accident Analysis and Prevention, 2020, 137:

[10]. Wang Li, He Mutian, Xu Shuo, Yuan Tian, Zhao Tianyi, Liu Jianfei. Garbage classification and detection based on yolov5s network[J]. Packaging engineering, 2021, 42(08): 50-56.

[11]. Henseler J, Ringle C M, Sarstedt M. A New Criterion for Assessing Discriminant Validity in Variance-based Structural Equation Modeling[J]. Journal of the Academy of Marketing Science, 2015, 43(1): 115-135.

[12]. Rezztofighi H, Tsoi N, Gwak J Y, et al. Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression[C]. Long Beach, USA,: IEEE Conference on Computer Vision and Pattern Recognition, 2019.