ISSN: 2472-3703 DOI: 10.6911/WSRJ.202201_8(1).0009

Driver Fatigue Detection Algorithm Based on Improved Yolov4

Yufa Liu, Ziwen Chen, Chao Xu, Ting Liu, Xianjiu Guo*

School of Information Engineering, Dalian Ocean University, Dalian, 116023, China

Abstract

Traditional driver fatigue detection methods include detection based on driver's physiological characteristics, detection based on motor vehicle behavior characteristics, and detection based on driver behavior characteristics. Now the analysis of driver's facial features is the main detection method. Commonly used face target detection methods include Faster R-CNN, MTCNN, and SSD; this article uses an improved algorithm for the YOLOv4 target detection method. Based on the basic framework of the YOLOv4 real-time target detection algorithm, the Focus framework, an advantage of YOLOv5, is added Into the Backbone structure of YOLOv4, and make corresponding improvements, and the CSP structure of YOLOv4 form a CSP+Focus structure. After a convolution operation of 32 convolution kernels, the target image is sliced. The detection of human faces shows that a certain detection speed and mAP value can be improved.

Keywords

Fatigue detection; YOLOv4, Focus structure; Target detection.

1. INTRODUCTION

In recent years, with the development of society and the progress of transportation, the frequency of traffic accidents has also increased greatly while people are convenient to travel. Among them, the instability of the driver's driving state leads to casualties in the annual accidents. It accounts for a large proportion and brings great threats and challenges to the lives and property of the people. According to statistics, in 2019, the number of deaths from traffic accidents in the country was 62,763, and there are about 1.24 million people worldwide each year. Every piece of data is extremely cruel and heavy. According to statistics from the World Health Organization, traffic accidents rank eighth among the accidents that cause human deaths of all ages. Therefore, it is very necessary to research and design a fatigue warning method for the driver. For the driver's fatigue driving, it mainly refers to the fact that the driver has been driving the vehicle for more than 4 hours without rest, physically and mentally showing eye fatigue, mental distraction, decreased concentration, increased blinking frequency and yawning frequency, and then the real-time road The accuracy of judgment of the situation is not high, leading to accidents and a series of tragic consequences. In response to this situation, the world's major automobile manufacturers and research institutions in various countries are researching and developing, and have developed some advanced models and algorithms. At present, the commonly used target detection algorithms include sliding window target detection; R-CNN, Fast-RCNN, and Faster-RCNN using Two-Stage; SSD and YOLO using One-Stage. The disadvantage of the sliding window algorithm is that the calculation volume is large. If the coarse-grained standard is used for frame selection, although the calculation amount can be reduced, but it will undoubtedly reduce the accuracy of the algorithm; the structure of the CNN series is mainly divided into convolutional layer and RPN layer, ROI layer and Classification layer[1]. Although the accuracy of this two-stage detection method based on the region-based detector is quite high, it has high performance requirements; therefore, this paper uses the

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202201_8(1).0009

improved algorithm of YOLOv4 to detect driver fatigue Verification experiment. Analyze the driver's face, compare the driver's different facial expressions under fatigue conditions and normal conditions, and summarize some typical fatigue characteristics. For example, the driver's head posture, blink frequency, line of sight direction and yawn detection, etc., detect and extract fatigue features from the driver's video image, calculate fatigue parameters, and finally determine the driver's fatigue level.

2. METHOD

2.1. Introduction to YOLOv4 Algorithm

The YOLO algorithm (You Only Look Once), as the name suggests, is that end-to-end target detection can be achieved by only performing a CNN operation. It is a real-time, one-stage target detection algorithm with high accuracy. The principle is to input the original image into the network and predict the category and position information of the object through loss calculations, and convert the complex target detection problem into a common regression problem. It was originally proposed by Joseph Redmon in 2016 and updated to the YOLOv3 version. At that time, the commonly used target detection algorithm in academia was based on the R-CNN series of Region Proposal algorithms. This series of algorithms worked well in instance segmentation and target detection, but compared to the later YOLO algorithm, its operation speed was relatively slow. , It is difficult to adapt to real-time detection tasks. The R-CNN series algorithm is Two-Stage, and its operation idea is to first generate the region for the target to be detected, and then solve the classification and regression separately [3]. The YOLO

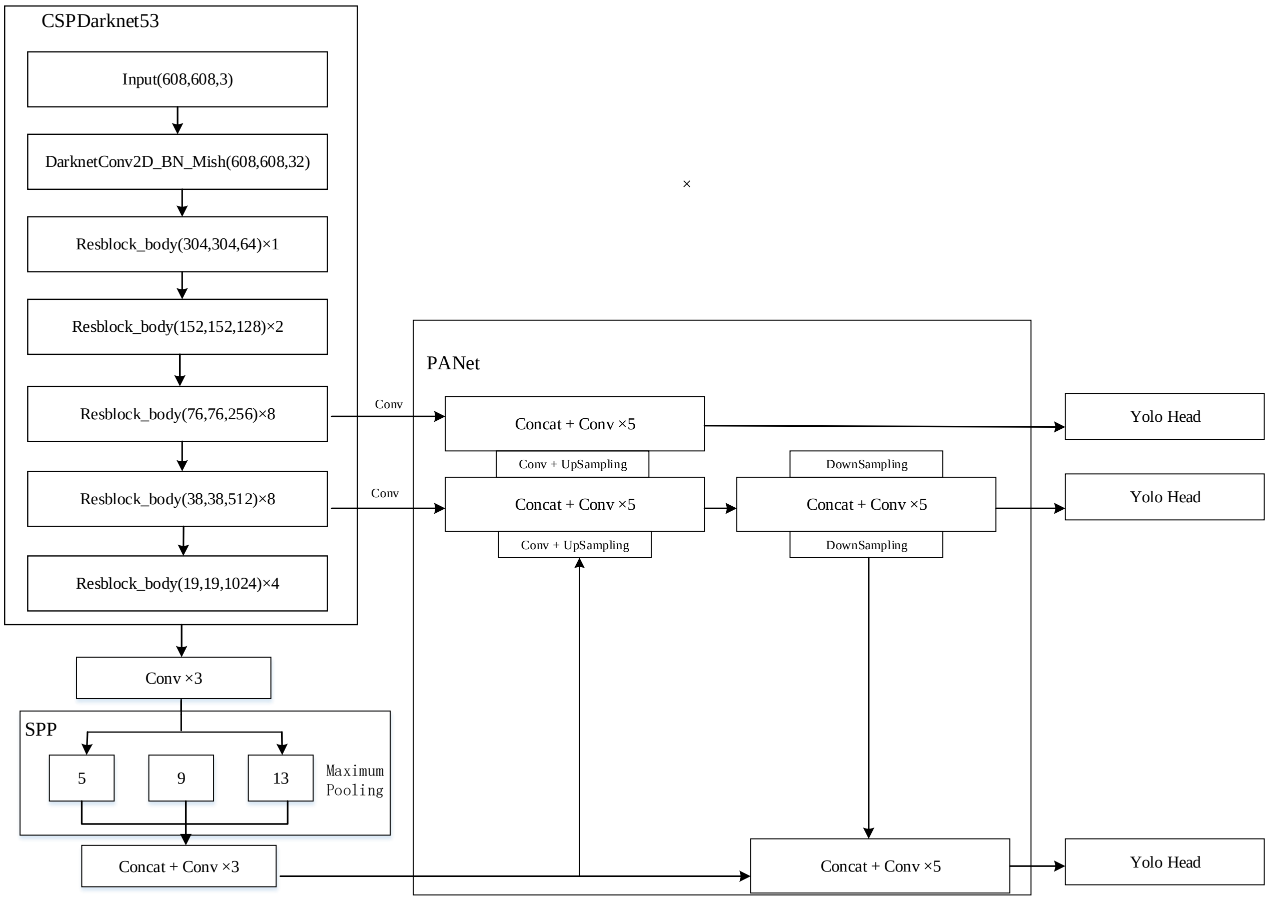

Figure 1. YOLOv4 network structure

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202201_8(1).0009

algorithm proposed by Joseph Redmon is One-Stage, when the target is detected It can be solved by skipping the region generation, outputting the location of the detection target and the category it belongs to, and using a single network, so it can be directly optimized end-to-end in the detection performance, so that the basic YOLO model can process the image in real time. The speed can reach 45 frames per second. The network structure of YOLOv4 is composed of Input, Backbone, Neck and Head [2]. The network structure of YOLOv4 is shown in Figure 1.

YOLOv4 has made some innovations on the basis of the first three versions, which are mainly reflected in the following 4 parts:

(1) Input: Improvements to the input during training, mainly including Mosaic data enhancement, cmBN, and SAT self-confrontation training

(2) Backbone: Integrate Tricks, which are not the latest excellent papers in the field of deep learning, into YOLOv4, such as CSPDarknet53, Mish activation function, and Dropblock, and the final effect is particularly good.

(3) Neck: The target detection network often inserts some layers between Backbone and the final output layer, such as SPP module, FPN+PAN structure

(4) Prediction: In the detection of overlapping targets, the effect of DIOU_nms is better than that of traditional nms.

2.2. Focus Structure

YOLOv4 includes four parts: Input, Backbone, Neck, and Prediction. For a single image, in addition to classic geometric distortion and illumination distortion, it also innovatively uses image occlusion (Random Erase, Cutout, Hide and Seek, Grid Mask, MixUp) Technology [9]. For the combination of multiple images, the author uses a mixture of CutMix and Mosaic technologies. And through Self-Adversarial Training (SAT) for data enhancement. YOLO extracts the features of the target to be detected through network features, and then uses the fully connected layer to obtain the predicted value. In the basic YOLOv4 network structure, the CSPNet structure is added, and the CSP structure is designed in Backbone. Makes YOLOv4 faster than the previous version, and extracts the features of the target to be detected faster, and the effect is better. In YOLOv5, the Focus structure is added. In this paper, the Focus structure in the YOLOv5 Backbone structure is added to the Backbone backbone network of YOLOv4 to form the Focus + CSP structure. The function of the Focus module in YOLOv5 is to slice the target image to be detected. As shown in Figure 2, the original image of 640 × 640 × 3 is input into the Focus structure, and through the slicing operation, it becomes 320 × 320 × 12 The feature map, after another convolution operation, finally becomes a 320 × 320 × 32 feature map. The specific operation is to get a value every other pixel in a picture, which is similar to adjacent down-sampling. In this way, four pictures are obtained [4]. The four pictures are complementary and similar in length, but no information is lost. The W and H information is concentrated in the channel space, and the input channel is expanded by 4 times. That is, the spliced picture has 12 channels compared to the original RGB three-channel mode. Finally, the processed image data is subjected to a convolution operation. Finally, a double down-sampled feature map without loss of information is obtained. By adding the Focus module by design, its own parameters and calculations have been improved. The calculation method of the parameter quantity and the count quantity of Focus structure is shown in Table 1.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202201_8(1).0009

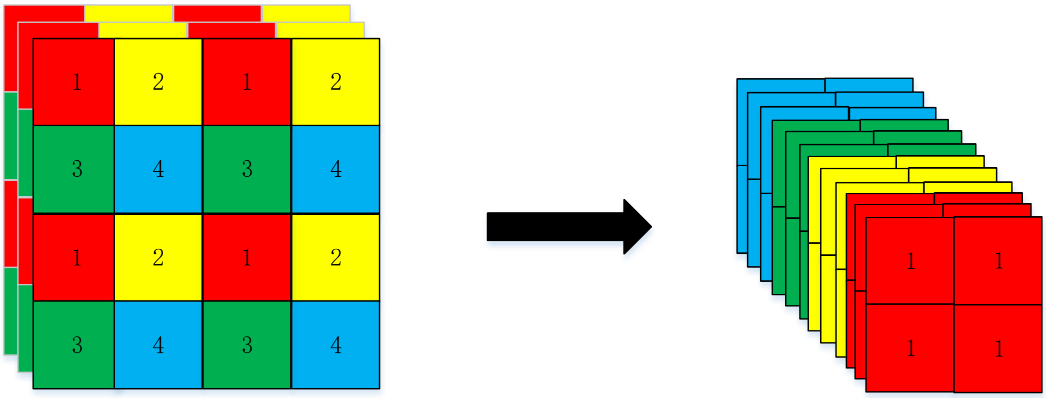

Figure 2. Focus structure

Table 1. Parameter and Count

Input	Kenel
Parameter	Kh × Kw × Cin × Cout
Count	params × H × W

3. EXPERIMENTS

3.1. Experiment Platform

The data set for the face fatigue detection experiment in this paper uses the seed-vig of Shanghai Jiaotong University. The data set contains EEG and EOG data. The fatigue evaluation standard uses the PERCLOS algorithm standard. The experimental platform parameters used for the experiment are shown in Table 2.

Table 2. Experiment Platform

Software And Hardware	Parameter
CPU	i7 11700
GPU	3080Ti
Video Memory	12G
Operating System	Windows10
Experiment Platform	Cuda11, Pytorch1.8

3.2. Experimental Result

The improved algorithm in this paper uses the indicators Precision, mAP@0.5, Recall, and F1 to quantify the experimental results to show the effectiveness of the experiment [5].

Precision, namely precision rate, can also be called precision rate, as shown in the calculation formula (1):

$$precision = \frac{TP}{TP+FP} \tag{1}$$

Where: TP is actually a positive sample and predicted as a positive sample; FP is actually a negative sample and predicted as a positive sample.

ISSN: 2472-3703
DOI: 10.6911/WSRJ.202201_8(1).0009

mAP, the average mean value. This indicator is a comprehensive measurement of the target to be detected. Common mAPs are represented by mAP@0.5.

Recall, recall rate, can also be called recall rate. It is the ability to evaluate the classifier to find all positive samples, and the number of positive samples correctly predicted to account for the total number of all positive samples, as shown in the calculation formula (2):

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

Where: TP is actually a positive sample and predicted as a positive sample; TN is actually a positive sample and predicted as a negative sample.

F1, a measure of classification, is actually the average of precision and recall.

Through the comparison of the experiment before and after the improvement, Precision, mAP@0.5, Recall, F1 have all achieved good results, and the improvement is obvious. The experimental results are shown in Figure 3 and Figure 4.

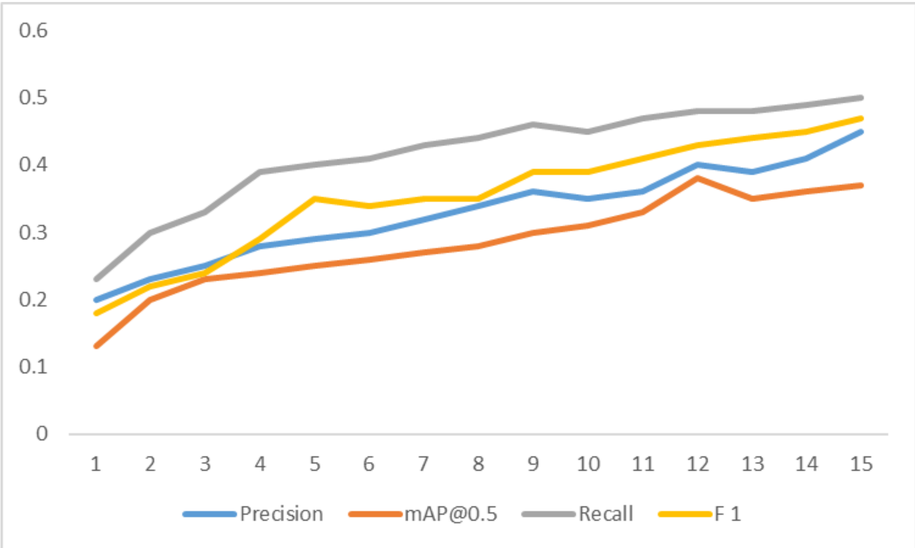

Figure 3. Before improvement

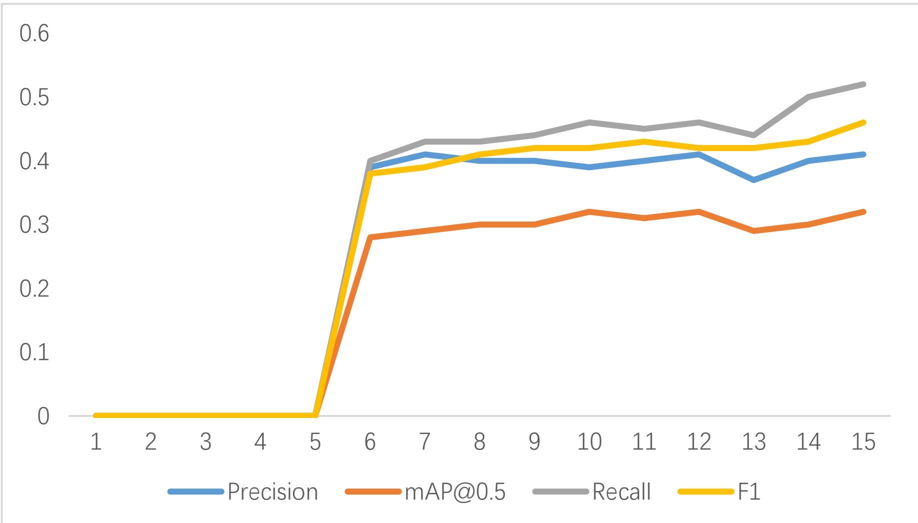

Figure 4. After improvement

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202201_8(1).0009

4. CONCLUSION

This paper uses the basic YOLOv4 target detection algorithm to detect the fatigue of the driver, and then introduces the Focus module in the Backbone structure of YOLO v5 to form a Focus+CSP structure, which improves the feature extraction ability of the network structure to be detected. Under the conditions of the experimental platform, Precision, Recall, mAP and other evaluation indicators have been improved, the effect is very obvious, and the detection speed and accuracy have been improved.

REFERENCES

[1] Liu Xue. Research on Fatigue Driving Detection Method Based on Face Image Analysis [D]. Nanjing University of Aeronautics and Astronautics, 2015.

[2] Zhang Luyao. YOLOv5s-based detection of whether a face is wearing a mask. Smart Computers and Applications 11.9 (2021): 9.

[3] Yu Weicong, Guo Xianjiu, Liu Yufa, et al. Sea treasure detection method based on lightweight deep learning Mobilenet-SSD network model[J]. Journal of Dalian Ocean University, 2021, 36(2):7.

[4] Li Zhaohui, and Zhang Weiliang. "Fatigue driving detection based on improved YOLOv4 algorithm." Electronic measurement technology 44.13(2021):6.

[5] Xu Zirui, Liu Meng, and Tan Yating. "Research on Vehicle Detection and Traffic Statistics Based on YOLOv4." Modern Information Technology 4.15(2020): 4.

[6] Hartley L, Horberry T, Mabbott N, et al. Review of fatigue detection and prediction technologies[J]. National Road Transport Commission, 2000: 1-41.

[7] Nguyen Quoc H, Truong Hoang V. Real-Time Human Ear Detection Based on the Joint of Yolo and RetinaFace[J]. Complexity, 2021, 2021.

[8] Li Ruoxi, Lu Xiao, Zhang Yuansheng, Li Yue, Liu Peng. Research on improved YOLOv4 algorithm for underground personnel detection[J]. Mining Research and Development, 2021, 41(11): 179-185. DOI: 10.13827/j. cnki.kyyk.2021.11.029.

[9] Zhang Luyao, Han Hua. Detection of whether a face wears a mask based on YOLOv5s[J]. Intelligent Computers and Applications, 2021, 11(09): 196-199.

[10] Zeng Ge. Research on Driver Fatigue State Detection Based on Deep Learning [D]. Hunan University, 2018.