ISSN 2616-5775 Vol. 5, Issue 3: 28-34, DOI: 10.25236/AJCIS.2022.050304

YOLOv5-based fatigue state detection method

Hekai Zhang[1], Sichao Cheng[2]

[1]School of Electronic Engineering, Heilongjiang University, Harbin, Heilongjiang, 150080, China
[2]School of Electronic Information and Automation, Tianjin University of Science and Technology, Tianjin, 300457, China

Abstract: *With the development of social economy and the continuous increase of car ownership in China, fatigue driving is one of the problems people need to focus on nowadays, however, most of the current fatigue state detection methods are affected by the problems of expensive, susceptible to the environment and complicated methods. In order to solve the above problems, this paper proposes a fatigue state detection method with YOLOv5m as the basic network model, which first enhances the original image and then improves the loss function. The experimental results show that the mean accuracy of this method can be as high as 95.6%, which is 4.13 percentage points higher than YOLOv4 and 6.2 percentage points higher than YOLOv3, and the model accuracy is as high as 98.27%, which is 4 and 5.2 percentage points higher than YOLOv4 and YOLOv3, respectively. The recall rate is 95.1%, which is 2 and 3.2 percentage points higher than YOLOv4 and YOLOv3, respectively. It proves the reliability and advantages of the method in this paper.*

Keywords: *fatigued driving, YOLOv5m, loss function.*

1. Introduction

According to the special research report on the market and investment strategy of China's traffic accident scene investigation and rescue equipment industry from 2019 to 2025, it can be concluded that the average number of traffic accidents is 190000 every year, of which 60000 people die from traffic accidents every year. Accidents are mostly caused by drivers' fatigue driving and other behaviors. According to the data of China Quality News Network, China's car ownership has reached 281 million and 418 million drivers in 2020. With the continuous increase of car ownership in China and the increasing number of drivers, the risk of frequent traffic accidents increases. The main causes of traffic accidents are fatigue driving, watching mobile phones and not looking at the road for a short time.

The world has been paying great attention to how to avoid fatigue driving, and more and more companies and research institutions have designed various Fatigue driving detection methods are generally divided into three methods based on physiological signals, based on Fatigue driving detection methods are generally divided into two methods based on physiological signals, based on vehicle information.

The physiological information detection method collects the driver's physiological signal through the sensor in direct contact with the human body to judge the driver's fatigue state during driving. The data obtained by this method is reliable, and the physiological signal comes directly from the human body. The fatigue state analysis is accurate and reliable. However, the device for detecting physiological signal should be installed on the driver, which has great interference to the driver.

The vehicle information detection method detects the driver's fatigue state according to the signal characteristics such as the driver's manipulation of the steering wheel, the change of rotation speed, whether the pressure on the accelerator pedal is stable, the driving track of the vehicle and so on. This detection method is simple, but it is easy to be affected by environmental factors, such as the good road environment, whether the climate is bad, and the anti-interference ability of the detection system is poor.

Along with the rise of convolutional neural networks, deep learning-based target detection methods have many wide [1-2] applications in artificial intelligence, information technology and other fields. The algorithms are mainly divided into two-stage (two-stage) and single-stage (one-stage) types. Two-stage is represented [3-5] by RCNN series, which has high accuracy but low timeliness. The one-stage type is represented by SSD [6]and YOLO [7-10] series, which have fast detection speed but slightly lower accuracy. The detection of driver fatigue status through new technologies such as machine vision has also received the attention of many related companies, among which RAMZAN M[11]et al. gave the latest research

ISSN 2616-5775 Vol. 5, Issue 3: 28-34, DOI: 10.25236/AJCIS.2022.050304

progress in this field, and Huang Jiacai [12] et al. gave a fatigue driving detection solution based on key points of the human face. On the industrial side, NOSD gives a machine learning approach to analyze driving behavior, which is comprehensive in analysis and variable in patterns, but vulnerable to the environment and costly.

In this paper, we use YOLOv5m network model for driver fatigue state detection, introduce the basic network model and optimize the detection algorithm, and compare it with YOLOv3 and YOLOv4. After validation, this algorithm has a good balance of accuracy and speed in driver fatigue detection, which proves the practicality of this method.

2. Introduction to the YOLOv5 network model

YOLOv5 can be structurally divided into 4 parts: Input, Backbone, Neck and Head of the backbone part, and its network structure is shown in Figure 1.

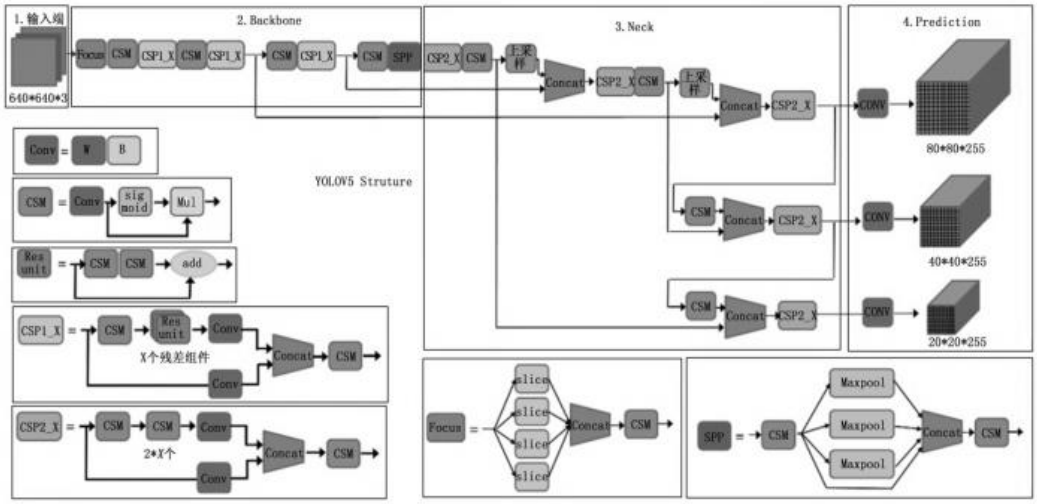

Figure 1: Network structure of YOLOv5

The input of YOLOv5 takes the input image and performs certain data processing, such as scaling to uniform size and then sends it to the network for learning. backbone contains Focus, CSP and SPP structures. csp1_X and csp2_X structures in CSP are used in Backbone and Neck respectively. SPP uses the maximum pooling 13, 9, 5, respectively, and then performs concat fusion to improve the perceptual field. These structures [13] can effectively avoid the problems of image distortion caused by cropping and scaling operations on image regions, and also solve the problem[14] of repeated feature extraction of images by convolutional neural networks. the Neck part adopts the structure of FPN and PAN combination to get a series of network layers that mix and combine image features, and pass the image features to the prediction layer, which enhances the information propagation and has the ability of preparing to retain spatial information. The Bounding box loss function at the output is GIOU_Loss, NMS performs non-maximum suppression on the final detection box of the target to obtain the optimal target box.

The activation functions used by yolov5 include leakyReLU and Sigmoid. yolov5 uses the leakyReLU activation function for the middle hidden layer and the Sigmoid activation function for the final detection layer. While yolov4 uses mish with the leakyReLU activation function, the backbone network uses mish. the complexity of the mish activation function is higher.

After comparing the performance of the four versions of YOLOv5, the faster and more accurate YOLOv5m is chosen as the base model.

3. Detection algorithm optimization

3.1. Data processing

In general target detection tasks, the image sizes of the data sets vary, and the common processing method is to scale the original images to a uniform standard size for both training and testing, and then

ISSN 2616-5775 Vol. 5, Issue 3: 28-34, DOI: 10.25236/AJCIS.2022.050304

feed them into the network [15]. However, because the actual detection is difficult due to the small target to be detected this time, the effect of using traditional data processing methods is not ideal, so the algorithm in this paper chooses to use data processing methods such as adaptive anchor frame calculation and adaptive image scaling at the input side.

Since the base anchor frames [116, 90, 156, 198, 373, 326], [30, 61, 62, 45, 59, 119], [10, 13, 16, 30, 33, 23] have been set before the network training, the network model will compare the predicted frames obtained from the training based on this anchor frame with the real frames, and update and iteratively adjust the network model parameters according to their differences in reverse[16].

The data processing of this algorithm is performed as follows: for training, the image is adaptively scaled and then scaled to a uniform size and fed into the network for training. For testing, only the adaptive image scaling is used, and the original image is adaptively scaled with a minimum of gray borders to reduce redundant information before being passed into the detection network to improve the inference speed during actual detection.

3.2. Improving the loss function

IoU [17] is called the intersection and merge ratio, which is an index to evaluate the effectiveness of target detection. It is calculated as the ratio of the intersection and merge between the prediction frame and the real frame, and GIoU [18] is a penalty term introduced on the IoU basis to facilitate a more accurate response to the intersection of the detection frame and the real frame. Both are calculated as follows.

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

$$GIoU = IoU - \frac{|C - (A \cup B)|}{|C|} \tag{2}$$

Where A denotes the detection box, B denotes the true box, and C represents the smallest external rectangular box containing the detection box and the true box, $|C - (A \cup B)|$ indicating the penalty term.

The yolo series loss calculation includes target confidence, category probability and bounding box regression loss. yolov5 uses GIoU Loss in the early stage of bounding box loss and CIoU Loss in the later stage, and CIoU Loss in yolov4, which brings faster convergence and better performance compared with other methods.As shown in Figs2.(a) and 2(b), GIoU can measure the distance and proximity of two frames when A and B do not intersect; as shown in Figs. 2(c) and 2(d), GIoU can also reflect the way of intersection of two frames. compared with IoU, GIoU can better distinguish the position relationship between the detected frame and the real frame. However, when C is equal to $A \cup B$. GIoU will degenerate into IoU, and then the GIoU advantage will disappear.

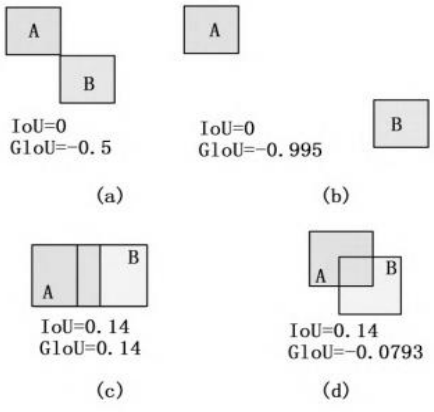

Figure 2: Location Relationship

Therefore, this paper choose CIoU_Loss [19] as the bounding box loss function to make the prediction box more closely match the real box. the CIoU calculation process is as follows.

$$CIoU = IoU - \frac{\rho^2\left(b, b^{gt}\right)}{c^2} - \alpha v \tag{3}$$

ISSN 2616-5775 Vol. 5, Issue 3: 28-34, DOI: 10.25236/AJCIS.2022.050304

Where, as shown in the figure 3, b and denote the centroids of the prediction frame and the real frame, b^{gt} respectively. $d = \rho\left(b, b^{gt}\right)$ denotes the distance between the center points of the two boxes, and c denotes the diagonal distance between the predicted box and the smallest outer rectangle of the real box.

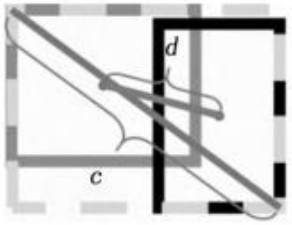

Figure 3: CIoU related diagrams

In Eq. (3), α is the parameter that does trade-off, v measuring the aspect ratio consistency, the formula is shown in Eqs. (4) and (5).

$$\alpha = \frac{v}{1 - IoU + v} \tag{4}$$

$$v = \frac{4}{\pi^2}\left(arctan\frac{\omega^{gt}}{h^{gt}} - arctan\frac{\omega}{h}\right)^2 \tag{5}$$

Where ω and ω^{gt} denote the width of the predicted border and the true border, respectively, and h and h^{gt} denote the height of the predicted border and the true border, respectively. the CIoU_Loss is calculated as

$$CIoU\_Loss = 1 - CIoU \tag{6}$$

4. Experiments and results

4.1. Data set

The experimental datasets are all collected from online, with a total of 4250 images, which provide a more comprehensive overview of the driver fatigue state. It was divided according to the ratio of 7:2:1 for training, validation and testing, with the number of images in the training set being 2975, the number of images in the validation set being 850 and the number of images in the testing set being 425. The data set is in PASVAL VOC format, and the images are annotated using the sprite tagging assistant, including closed_eye, closed_mouth, open_eye, _open_mouth, where closed_eye means the person to be detected is closing his eyes, closed_mouth means he is not opening his mouth, open_mouth means he is opening his mouth, and open_eye means he is opening his eyes.An example of the data set is shown in the figure 4. The number of markers is shown in the figure 5.

Figure 4: Example data set

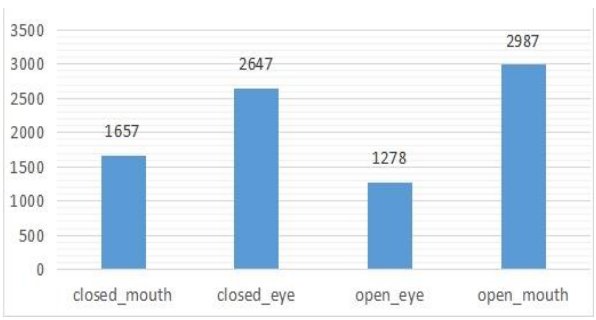

Figure 5: Classification of dataset categories

4.2. Experimental environment and model training

The configuration of this experiment is shown in the table.1

ISSN 2616-5775 Vol. 5, Issue 3: 28-34, DOI: 10.25236/AJCIS.2022.050304

Table 1: Experimental environment configuration

Parameters	Configuration
System Environment	Windows10
CPU	Intel(R) Core(TM) i5-8250U CPU @ 1.60GHz 1.80 GHz
GPU	GeForce MX150
GPU acceleration	CUDA11.4
Training framework	PyTorch
Language	Python3.8

The training parameters are set as follows: input image size is 640*640, learning rate is set 0.02, learning rate period is set 0.25, training batch size is 44 set, total training 50 epochs, batch-size is set 1.

4.3. Analysis of experimental results

From Equation (7), the loss value of YOLOv5 consists of three categories of losses, where box_loss indicates object position loss, cls_loss indicates object category loss, and obj_loss indicates whether the loss contains the target object.

$$Loss = box\_loss + cls\_loss + obj\_loss \tag{7}$$

The following performance metrics[20] are used to evaluate the performance of the algorithm in this paper: accuracy (P, precision), recall (R, recall), and mean average precision (mAP, mean average precision).

Accuracy is a measure of precision and indicates the proportion of examples classified as positive that are actually positive examples. Recall is a measure of coverage and counts the number of positive examples that are classified as positive. The formula is as follows.

$$Precision = TP/(TP + FP) * 100\% \tag{8}$$

$$Recall = TP/(TP + FN) * 100\% \tag{9}$$

Where: TP represents the number of predicted positive cases that are also actual positive cases, FP represents the number of predicted positive cases but actual negative cases, and FN represents the number of predicted negative cases but actual positive cases.

The mAP is used to measure the recognition accuracy, which is obtained by averaging the AP values of all categories. In which the P-R curve can be plotted by calculating the highest accuracy rate at different recall rates, and the area enclosed by this curve is the AP value of that category. The highest Precision, Recall, mAP_0.5, mAP_0.5:0.95 of the trained model can reach 0.9827, 0.95113, 0.95621, respectively 0.8668, as shown in Fig.6.

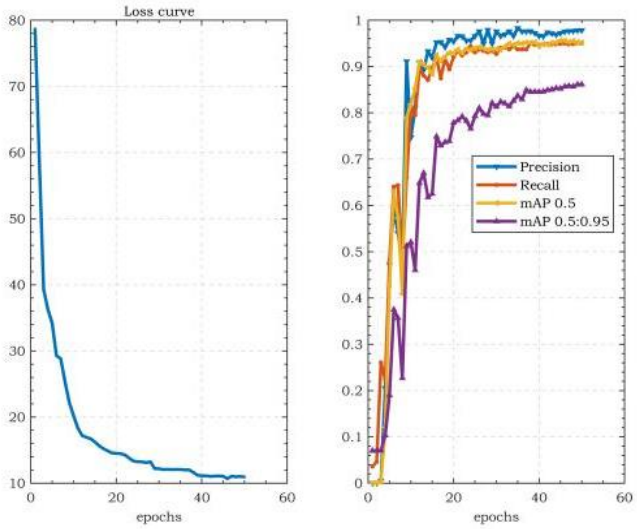

Figure 6: Model performance

ISSN 2616-5775 Vol. 5, Issue 3: 28-34, DOI: 10.25236/AJCIS.2022.050304

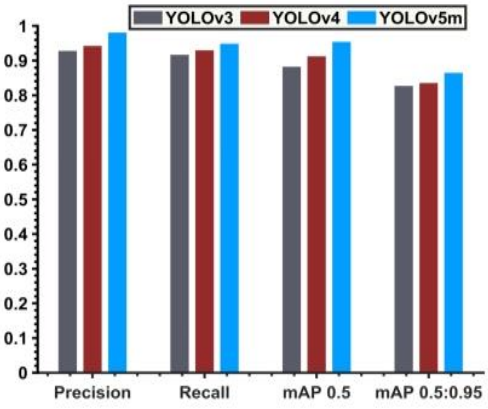

Figure 7: Comparison of the evaluation index performance of different models

To verify the effectiveness of the algorithm in this paper, this method was compared with YOLOv4 and YOLOv3, and the comparison results are shown in Fig 7. It can be concluded that the fatigue state detection method using YOLOv5m as the network model has outstanding advantages in terms of recall, accuracy and average accuracy mean, which proves the effectiveness of this paper's method applied to fatigue state detection. Meanwhile, in order to further verify the effectiveness and capability of this paper using the model for fatigue state detection, the pictures are sent into the network model for testing after adaptive deflation and data enhancement, and the results are obtained in only 0.047 seconds as 8 shown in Fig.

Figure 8: Fatigue state detection results

5. Conclusion

In summary, this paper proposes a fatigue detection algorithm with YOLOv5m as the basic network model in order to balance the speed and accuracy of detecting driver fatigue driving status, and performs data enhancement processing while improving the team loss function to improve the detection accuracy. The comparison and validation with other algorithms show that the average accuracy of this method is 95.6%, the model accuracy is 98.27%, and the recall rate is 95.1%, all of which have some improvement over YOLOv3 and YOLOv4, proving that this method has some advantages. However, this method still has many shortcomings, and the fatigue state detection is relatively weak when dealing with the obscured target state, and there will be a certain degree of missed detection. Therefore, further exploration will be conducted in this aspect next.

References

[1] Zhang Hui, Wang Kunfeng, Wang Feiyue. Progress and Prospects of Deep Learning in Target Vision Detection [J]. Journal of AutomationJournal, 2017, 43(08): 1289-1305. DOI: 10.16 383/

[2] Ma Qifeng. Research on military target detection technology based on deep learning [D]. North Central University, 2020. DOI: 10.27470/

[3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Region-based convolutional networks for accurate object detection and segmentation [J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 38(1): 142-158.

ISSN 2616-5775 Vol. 5, Issue 3: 28-34, DOI: 10.25236/AJCIS.2022.050304

[4] GIRSHICK R, Fast R-CNN [C]//Proceedings of the IEEE international conference on computer vision, 2015: 1440-1448,
[5] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks [J]. Advance in neural information processing systems, 2015, 28: 91-99.
[6] LIU W, ANGUELOV D, ERHAN D, et al, SSD: Single shot multibox detector [C] //European conference on computer vision, Springer, Cham, 2016: 21-37.
[7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Reconition, 2016: 779-788.
[8] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.
[9] REDMON J, FARHADI A, Yolov3:an incremental improvement [C] //IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2018, 87(8): 101-104.
[10] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: optional speed and accuracy of object detection [C] //IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2020, 57(5): 9-12
[11] RAMZAN M, HIKMAT U K, SHAHID A, et al. A survey on State-of-the-Art drowsiness detection techniques [J]. IEEE Access, 2019(7): 61904-61919.
[12] Huang J. C., Kuang W. T., Mao Kuan-Cheng. Research on fatigue driving detection based on key points of human face [J]. Journal of Nanjing Engineering College (Natural Science Edition), 2017, 15(04): 8-13. DOI: 10.13960.
[13] Yu Shuo, Li Hui, Gui Fangjun, Yang Yanqi, Lv Chenyang. Real-time detection algorithm of mask wearing based on YOLOv5 in complex scenes [J]. Computer Measurement and Control, 2021, 29(12): 188-194. DOI: 10.16526.
[14] Huang, Lin-Quan, Jiang, Liang-Wei, Gao, Xiao-Feng. Improved real-time video helmet wearing detection algorithm for YOLOv3 [J]. Modern Computing Computer, 2020(30): 32-38+43.
[15] ZHANG S J, GAO R, Research on Visual Image Processing of Mobile Robot Based on OpenCV [J]. Journal of Computer Science, 2017, 28(5): 255-275.
[16] Tan S.L., Beixiongbo, Lu Gonglin, Tan Xiaohu. Real-time detection of personnel mask wear based on YOLOv5 network model [J]. Laser Miscellany DOI:10.14016.
[17] JIANG B, LUO R, MAO J, et al. Acquisition of localization confidence for accurate object detection [C] //Proceedings of the European conference on computer vision (ECCV), 2018: 784-799.
[18] REZATOFIGHI H, TOSI N, GWAK J Y, et al, Generalized intersection over union: A metric and a loss for bounding box regression [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 658-666.
[19] ZHENG Z, WANG P, REN D, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation [J]. IEEE Transactions on Cybernetics, 2021,26(8):1-13.
[20] FOMIN I S, BAKHSHIEV A V, GROMOSHINSKII D A. Study of using deep learning nets for mark detection in space docking control images [J]. Procedia Computer Science, 2017, 103: 59-66.