Traffic signal image detection technology based on YOLO

Hui Gao, Wenhao Wang*, Chengjin Yang, Weile Jiao, Ziwei Chen and Tong Zhang
Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Huai’an, Jiangsu, 223003, China
*Corresponding author’s e-mail: wangwenhao@hyit.edu.cn

Abstract. The detection and recognition of traffic signal image is an important content in intelligent transportation system. It can be applied to driver assistance system to effectively recognize traffic signal signs on the road, so as to reduce the occurrence of traffic accidents. At the same time, it also provides strong technical support for the future unmanned driving system. The main content of this paper is based on the deep learning method, using YOLOv3 and YOLOv4 algorithm to detect and recognize the traffic signal image on the road. The experimental results show that the recognition result of YOLOv4 network is better than that of YOLOv3 network.

1. Introduction
Traffic signal image is one of the most important information in the road. The research of traffic signal image detection and recognition has made great progress, but the traditional detection algorithm has defects in the accuracy and speed of detection. Common target detection algorithms can be divided into two categories [1], for example, R-CNN, Fast R-CNN, Faster R-CNN and other algorithms, which need to form a candidate box containing the target to be detected in the detection process; On the other hand, without such a candidate box, we can directly extract features and detect targets, such as YOLOv1, YOLOv2, YOLOv3, YOLOv4, SSD and RetinaNet. YOLO algorithm has better detection accuracy and speed, and better comprehensive performance.

Reference [2] focuses on the detection and recognition of small targets. It improves the yolov4 algorithm by K-means clustering the prior box on the multi-scale branch graph, and then speeds up the feature extraction speed by cutting the network branches and compressing the convolution layer. It also improves the accuracy of the prediction box boundary by improving the loss function. Thus, the recognition accuracy in small target recognition task is improved.

In reference [3], a single-stage deep neural network (df-yolov3) is proposed to improve the traditional yolov3 algorithm, considering the serious positioning error of YOLO algorithm. The target features are extracted by enhancing the deep residual network, and three convolution feature maps of different scales are added to fuse the feature maps of corresponding scales in the residual network, the final feature pyramid is formed to perform the target prediction task.

The above methods provide a reference method for this paper. In this paper, the traffic signal image is detected and recognized by using the network structure of yolov3[4] and yolov4[5] respectively, and the detection results of the two network models are compared.
2. Introduction of YOLO algorithm

2.1. YOLO3
YOLOv3 divides the given image into S×S grids, each grid is only responsible for the target in which the center falls, and each grid will predict B bounding boxes and their corresponding confidence. Each bounding box consists of five parameters \((x, y, w, h, c)\), where \((x, y)\) represents the location of the bounding box; \(w, h\) is the ratio of the width and height of the bounding box to the whole image, which can be called the width and height of the bounding box; \(c\) value is the confidence of the boundary box, including the probability of the existence of the target in the boundary box and the accuracy of prediction. The calculation formula of standardization is as follows:

\[
\hat{x}^k = \frac{x^k - E\left[\hat{x}^k\right]}{\sqrt{\text{Var}\left[\hat{x}^k\right]}}
\]

YOLOv3 uses the scheme of feature pyramid, which uses three different scale feature layers to convolute for five times to predict the feature image. Through multi-scale fusion, the detection accuracy of small objects has been greatly improved.

2.2. YOLO4
The network structure of YOLOv4 algorithm is more complex than that of YOLOv3, which is an improved version of YOLOv3 algorithm. Based on YOLOv3, the main feature extraction network, data enhancement, activation function and loss function are optimized, so as to improve the detection speed and accuracy. The original activation function of DarknetConv2D in YOLOv3 is modified from LeakyReLU to Mish activation function. The function formula is as follows:

\[
Mish = x \times \tanh\left(\ln\left(1 + e^x\right)\right)
\]

The convolution block is constructed by DarknetConv2D_BN_Leaky changed to DarknetConv2D_BN_Mish, using the CSPNet structure, adds CSP to each large residual block of darknet53 to split the feature mapping of the base layer. CSPNet can improve CNN's learning ability, keep the accuracy and reduce the cost of calculation. In addition, YOLOv4 uses the structure of PANet in three effective feature layers. PANet is an instance segmentation algorithm. Its main feature is to extract features repeatedly, first from the bottom to the top, and then from the top to the bottom, so as to improve the detection effect of small objects.

3. System design

3.1. System architecture
The research background of this project is intelligent vehicle automatic driving technology, which can accurately detect signs through vehicle detection and recognition of traffic signal image.

The overall model of traffic signal image detection and recognition system is divided into the following three parts: data acquisition, data preprocessing and experiment. Compared with image recognition, the problem of traffic signal image recognition is more difficult. We need to recognize many types of traffic signal image, and also can achieve the effect of real-time detection to get the accurate location of the target.

3.2. Image data preprocessing
In the process of taking sample video, there are some problems, such as shooting angle, insufficient illumination, visibility degradation caused by bad weather. The quality of the obtained image has a certain degree of decline. In order to improve the image quality, in the target detection, image enhancement technology is often used to improve the clarity of the image, so that the target features in the image are more obvious.
(1) Algorithmic thinking

Contrast enhancement algorithm based on exposure fusion framework in order to make the image of all pixels well exposed, this method enhances the illumination in the dark area of the image, weakens the illumination in the over exposed area, and has little influence on the well illuminated area:

\[ R_i = \sum_{i=1}^{n} W_i P_i \]  

Where \( n \) is the number of images, \( P_i \) is the \( i \)th image in the exposure set, \( W_i \) is the weight map of the \( i \)th image, \( C \) is the index of the three color channels, and \( R \) is the enhanced result. The three color components are equal, and all the pixels are uneven: the pixel with good exposure has a larger weight, and the pixel with poor exposure has a smaller weight. The weights have been standardized. Therefore \( \sum_{i=1}^{n} W_i = 1 \). The low brightness image enhancement algorithm based on the camera response model can obtain the conversion formula between images with different exposure rates in the same scene:

\[ P_e = g(P, K) \]  

\[ g(P, K) = \beta P^\gamma = e^{(1-\gamma)P^\beta} \]  

Among \( \beta \) and \( \gamma \) They are two model parameters, which can be calculated according to camera parameter \( A, B \) and exposure rate \( K \). In order to reduce the computational complexity, the input image of two kinds of exposure fusion is taken as an example:

\[ R_e = W \times P_e + (1-W) \times g(P_e, K) \]  

(2) Image optimization and comparison

In order to show the effect of image enhancement algorithm, we select a typical image.

![Figure 1 after a rainy day](image)

From the enhanced image, we can see that for the image with low brightness of the whole image (Figure 1), after the image enhancement processing; It can be seen that the image enhancement method used in this paper has a good effect.

3.3. Data set construction of training network

The data set used in the training network is TT100K data set and 100 images captured in the car video. For some small number of traffic signs, the data enhancement method is used to expand the image. The final data set contains 10267 images, a total of 182 categories. The image size is 2048 * 2048, the image format is JPG, and the image background is traffic road, the foreground of the image is the traffic target. Among them, 90% were selected as training set and 10% as test set.

4. Experiment and analysis

4.1. Experimental environment

In order to verify the detection effect, this paper uses a computer with 8G memory, NVIDIA GeForce GTX 1650 independent graphics card, cuda11.0 acceleration, Python development language and python deep learning framework under Windows 10 operating system.
4.2. Evaluation criterion

4.2.1. Interaction ratio
During the experiment, the accuracy of prediction box position is evaluated by calculating the value of Intersection over Union. The IOU is the ratio of the predicted border to the real border, as shown in Figure 2.

![Figure 2 intersection and union ratio IOU](image)

Generally, as long as the IOU value is greater than or equal to 0.5, it is judged as the correct test result, as shown in the following formula.

\[
IOU = \frac{\text{area}(A \cap B)}{\text{area}(A \cup B)} \geq 0.5
\]

(7)

Where a is the predicted border of the model and B is the real border of the target. In the ideal case, the value of IOU is 1, which indicates that the performance of the detected model is good, and the real border completely coincides with the predicted border.

4.2.2. Accuracy rate and call rate
In order to accurately evaluate the performance of the target detection algorithm, it is necessary to calculate the Precision and Recall of the detected target. Accuracy refers to the proportion of the number of correct targets in all the detected targets after detection. Recall rate refers to the proportion of all positive samples in the whole dataset with the correct number of targets.

\[
\text{Precision} = \frac{TP}{TP + FP}
\]

(8)

\[
\text{Recall} = \frac{TP}{TP + FN}
\]

(9)

In the above formula, TP represents the same number of results detected by the model as the real ones; FP represents the number of targets that are inconsistent with the real results; FN indicates the number of targets that should have been detected by the model but missed.

4.2.3. Loss function
The loss function of YOLO consists of three parts: coordinate error, IOU error and classification error. The model parameters are optimized by calculating the mean square sum error of S*S*(B * 5+C) vector of network output and S*S*(B * 5+C) vector of real target input. The loss function is as follows:

\[
\text{loss} = \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \sum_{l=0}^{1} \hat{1}_{ij} \left[ (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\
+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \sum_{l=0}^{1} \hat{1}_{ij} \left[ (\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \right] \\
+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \sum_{l=0}^{1} \hat{1}_{ij} \left( C_i - \hat{C}_i \right)^2 + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \sum_{l=0}^{1} \hat{1}_{ij} \left( C_i - \hat{C}_i \right)^2 \\
+ \sum_{i=0}^{S^2} \sum_{l=0}^{1} \hat{1}_{ij} \sum_{c=\text{class}} \left( p_l(c) - \hat{p}_l(c) \right)^2
\]

(10)

The first two parts of the formula calculate the coordinate error, the middle two parts calculate the IOU error, and the last part calculates the classification error. Where \( \hat{1}_{ij} \) indicates that if the target
appears in cell I, it is 1; otherwise, it is 0; $1_{iy}^b$ and $1_{iy}^{no}$ indicate that the target appears or does not appear in the j border of the I cell, respectively.

4.3. Experimental results and analysis

4.3.1. Compare with each other

In this experiment, we train the prepared traffic sign data set by using two algorithms: YOLOv3 and YOLOv4.

Some of the detection results are shown in Figure 3. The left side is the detection result of YOLOv3, and the right side is the detection result of YOLOv4. The closer the IOU value is to 1, the higher the coincidence degree between the predicted border and the real border is, and the closer the detection model is to the expectation. It can be seen from the figure that the accuracy of the prediction box position is evaluated by comparing the values of the intersection and union ratio. The detection model of YOLOv4 has good performance, and the real frame basically coincides with the predicted frame. In such a coincidence degree, if the accuracy rate and recall rate can reach a very high value, this is the goal we expect to achieve.

4.3.2. Accuracy rate and recall rate

Selecting representative categories, we can see that the trained network model of YOLOv3 and yolov4 can basically achieve more than 75% recall rate of traffic signs on the test set, and the accuracy rate can reach 80%. By comprehensive comparison, it is found that both YOLOv3 and YOLOv4 have higher accuracy and recall.

According to the training, different loss curves of the two algorithm models are obtained, as shown in Figure 4. The left figure shows the loss curve of YOLOv3 algorithm model when training traffic signs, and the right figure shows the loss curve of YOLOv4 algorithm model when training traffic signs:
Figure 4 Loss comparison of two algorithms

Through the figure above, we find that the loss curves of the two algorithm models show a downward trend, which is caused by the back propagation of the deep neural network. In the process of repeated training, the error decreases continuously, and the loss value decreases continuously. However, comparing the curves of the two algorithms, we can clearly see whether it is the fluctuation amplitude of loss or the final AVG. In the numerical simulation of loss, the algorithm model of YOLOv4 is better, and its loss can achieve a better convergence effect in the training process.

5. Conclusions
In this paper, we use the YOLOv3 and yolov4 network to detect and recognize the traffic signal image. Firstly, we introduce two kinds of YOLO algorithms, followed by the collection and production of data sets and the distribution of samples. Secondly, we introduce the evaluation criteria of the experiment, and evaluate the two kinds of YOLO algorithms. Finally, we get the results of this experiment: in the aspect of detection target, we can get the following conclusions, the detection effect of the two kinds of YOLO algorithm is almost the same. In the recognition accuracy, the average accuracy of YOLOv4 algorithm is higher than that of YOLOv3 algorithm, that is, the recognition effect is good.

The research of traffic signal image detection and recognition is a very meaningful and challenging research, because considering various reasons, such as weather, visibility, temperature and other external factors will interfere with the detection. With the progress of science and technology, the progress and improvement of detection algorithm, we can reduce the interference of these external factors to the minimum. Science is always in progress, we are always in progress, I believe that in the future scientific development road, this technology will become better and better, and become an important step in the scientific and technological revolution.

Acknowledgments
This work was supported in part by the Industry university research cooperation project in Jiangsu Province, China (No. BY2020068), Qing Lan Project of JiangSu Province, Opening Project of Henan Engineering Laboratory of Photoelectric Sensor and Intelligent Measurement and Control ( Henan Polytechnic University)(No. HELPSIMC-2020-002). The authors would like to thank the anonymous reviewers and the editors for their many helpful suggestions.

References
[1] Shilei, B., Kexin, Y., Jianqi, Z. (2020) Lightweight YOLOv3 Traffic Sign Detection Algorithm. JISUANJI YU XIANDAIHUA, 09: 83-88.
[2] Xiaofeng, T., Zhijie, W. (2020) Table tennis recognition based on improved algorithm of yolov4. Technology Innovation and Application, 27: 74-76.
[3] Fukai, Z., Feng, Y., Ce, L. (2019) Fast Vehicle Detection Method Based on Improved YOLOv3. Computer Engineering and Applications, 55: 12-20.

[4] Xusheng, S., Lin, Z., Zhengtong, Y. (2020) A Vehicle Recognition Algorithm Based on Deep Learning. Heavy truck, 06: 40-41.

[5] Junlin, G., Xin, Z. (2020) Mask Wearing Detection Method Based on YOLOv4 Convolutional Neural Network. Modern Information Technology, 04: 9-11.