Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 8826593, 10 pages
https://doi.org/10.1155/2021/8826593

Research Article

Detecting Small Chinese Traffic Signs via Improved YOLOv3 Method

Baojun Zhang,[1] Guili Wang,[1,2] Huilan Wang,[1] Chenchen Xu,[1] Yu Li,[1] and Lin Xu[3]

[1]*School of Physics and Electronic Information, Anhui Normal University, Wuhu 241002, China*
[2]*Anhui Provincial Engineering Laboratory on Information Fusion and Control of Intelligent Robot, Wuhu, Anhui 241002, China*
[3]*School of Mathematics and Statistics, Anhui Normal University, Wuhu 241002, China*

Correspondence should be addressed to Guili Wang; xlpyhwgl@ahnu.edu.cn

Received 25 September 2020; Revised 13 January 2021; Accepted 22 January 2021; Published 3 February 2021

Academic Editor: Weijun Zhou

Copyright © 2021 Baojun Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Long-distance detection of traffic signs provides drivers with more reaction time, which is an effective technique to reduce the probability of sudden accidents. It is recognized that the imaging size of far traffic signs is decreasing with distance. Such a fact imposes much challenge on long-distance detection. Aiming to enhance the recognition rate of long-distance small targets, we design a four-scale detection structure based on the three-scale detection structure of YOLOv3 network. In order to reduce the occlusion effects of similar objects, NMS is replaced by soft-NMS. In addition, the datasets are trained and the K-Means method is used to generate the appropriate anchor boxes, so as to speed up the network computing. By using these methods, better experimental results for the recognition of long-distance traffic signs have been obtained. The recognition rate is 43.8 frames per second (FPS), and the recognition accuracy is improved to 98.8%, which is much better than the original YOLOv3.

1. Introduction

In recent years, the research on intelligent driving has attracted increasing attention. An important topic in this field is the detection of traffic signs. This is due to the fact that traffic signs are key facilities guiding smooth passage of vehicles, pedestrians, and regulating traffic behavior [1].

Currently, the traffic sign recognition mainly relies on the extraction of color features, shape features, and other methods [2]. The image is segmented according to various colors. Dai et al. proposed a new solution to improve the recognition rate of traffic signs in different brightness environments via colors, providing 78% accuracy and 11 FPS, respectively [3]. In [4], Miao used an improved K-Means clustering algorithm to segment color images and then used Hough transform to segment traffic signs of different shapes. The accuracy of this method is high, but the FPS is only 4.9. Among the above methods, the highest speed is 11 FPS and the lowest is 1.7 FPS. It is obvious that the speed cannot meet the needs of rapid detection, especially for small targets [5]. Therefore, machine learning methods have recently been introduced into this field, including SPPnet [6], R-CNN [7], Fast R-CNN [8], and Faster R-CNN [9]. In [10], Yao et al. proposed a traffic sign recognition method using histogram of oriented gradient-support vector machine (HOG-SVM) and grid search (GS) to detect traffic signs. The highest mAP of this method is 97.52%. In [11], Zuo et al. used Faster R-CNN to detect traffic signs with a map of 34.49%. In [12], Song et al. proposed a constitutional neural network with a small number of parameters to achieve the detection of traffic signs, and its mAP is 88%.

Research to improve the speed and accuracy of identifying targets is always on the way. Recently, several rapid detection methods based on regression, such as YOLO [13], YOLOv2 [14], and YOLOv3 [15], were developed. YOLOv3 proposed by Redmon can divide the image into 7×7 grids. The grid containing the center point of the target is responsible for predicting the target [16]. The YOLO method produces an accuracy of 64% for people, bicycles, and cars in traffic scenes. To improve the detection efficiency of YOLO on small targets, YOLOv2 is proposed by setting Darknet-19 as its backbone, while it uses K-Means [17] for clustering to

generate anchor boxes. A new method is proposed in YOLOv2, in which details of low-level features are introduced into high-level features. By using this strategy, the high-level feature graph contains more comprehensive information. Garg et al. used YOLOV2 to detect traffic signs, and its mAP is 77.89% [18]. A further improvement of YOLOv2 is YOLOv3, which is a fast target detection method with Darknet-53 [15]. The computational capacity by Darknet-53 is almost 1.5 times that of the one in Resnet-101 and twice the one in Resnet-152 [15]. Khan et al. proposed a new algorithm to improve the detection rate of traffic signs by using YOLOv3 and image enhancement [19]. The resulted mAP is 98.15% on KTSD and the FPS is 15.9.

As we all know, in intelligent transportation, quickly identifying long-distance traffic signs and fast alarming to drivers can help drivers to handle emergencies with more time. In this paper, we focus on the method of quick recognizing of Chinese traffic signs. There are two factors that affect the recognition of traffic signs by driving vehicles: driving speed and icon size. The size of the image is inversely proportional to the real distance from the target to the camera lens. However, when the traffic sign is too far from the camera, there will be many instances of false detection and missed detection when using YOLOv3. Therefore, we intend to increase the detection layer to YOLOv3 to improve the detection accuracy of small traffic sign images. In addition, to solve the problem that traffic signs are not easily recognized when they are blocked, we replace NMS (c.f. [20]) by soft-NMS (c.f. [21]). Furthermore, NMS (c.f. [20]) will be replaced by soft-NMS (c.f. [21]) to improve the recognition rate of traffic signs blocked. After improved YOLOv3 is trained with dataset, it is found that it has higher detection accuracy for long-distance traffic signs.

2. A Brief Instruction to YOLOv3

As shown in Figure 1, YOLOv3 consists of four parts: the picture preprocessing, the Darknet-53 feature extraction, the multi-scale detection, and the output layer. The image is preprocessed to become 416×416 size, and then it enters the feature extraction layer modeled on the Darknet-53. The obtained feature map enters the multi-scale detection network, and three-scale detection layers are extracted. Finally, the best bounding box is output by the output layer of YOLOv3. The image feature information is extracted by YOLOv3 modeling on the Darknet-53 and the multi-scale detection structure, which is superior to the detection accuracy of YOLO and YOLOv2 for small targets.

From the 1st layer to the 74th layer in YOLOv3, the number of convolution layers is 53, and the rest of the layers are residual layers [22]. We illustrate the detailed structure of the Darknet-53 with the feature map of 208×208 size, shown in Figure 1, where "$\times$ 2" indicates that there are two identical modules connected. The last layer of each size structure represents the feature map obtained by the residual structure. We mark it with purple. And the residual structure can avoid the gradient disappearance in the deep convolution [23]. The detailed residual structure is shown in Figure 2, which is a feature map of 208×208 residual structure, where x represents the information of the first 208×208 feature map. The image preprocessed is convoluted as the first 208×208 feature map x, which is filtered by 1×1 convolution kernel to obtain the second 208×208 feature map. Then, the second 208×208 feature map is padded, which is filtered by 3×3 convolution kernel to get the third 208×208 feature map. Finally, the fourth 208×208 feature map with information $H(x)$ is derived from $F(x)$. The connection mode is shown in Figure 2, where $H(x)$ is specified by

$$H(x) = x + F(x). \tag{1}$$

The advantage of this configuration is that when input $F(x)$ is small, $H(x)$ still has a large value, which can effectively avoid the gradient disappearing. Thus, the image can be more fully extracted features by adding residual structure in the feature structure, so the target in the image can be accurately identified and located.

3. Method of This Paper

3.1. Improved Multi-Scale Detection Network. The image is detected by multi-scale detection network when the feature information is extracted. However, multi-scale network is not better for small targets, so that the detection accuracy of small targets is weaker than that of large targets due to the small amount of information. Therefore, we propose an improved traffic sign detection method via YOLOv3, which can accurately identify small targets at a long distance. YOLOv3 has three different prediction boxes: 13×13, 26×26, and 52×52. It extracts image features from three scales, which is similar to the feature pyramid network.

We divide the traffic signs smaller than 32×32 pixels into small targets according to the evaluation index of Microsoft COCO [24]. After feature extraction by Darknet-53 in YOLOv3, the image enters the multi-scale detection layer, where the target in the original image is scaled down. For example, if the original image is 1024×1024, the small target pixels in the 13×13 feature map are given by

$$N_{13} \leq \left(32 \times \frac{13}{1024}\right) \times \left(32 \times \frac{13}{1024}\right) < 0.2. \tag{2}$$

Similarly, the number of pixels of small target in the feature maps of 26×26 and 52×52 is

$$\begin{aligned} N_{26} &\leq \left(32 \times \frac{26}{1024}\right) \times \left(32 \times \frac{26}{1024}\right) < 0.67, \\ N_{52} &\leq \left(32 \times \frac{52}{1024}\right) \times \left(32 \times \frac{52}{1024}\right) < 2.65. \end{aligned} \tag{3}$$

In summary, even in the 52×52 detection layer, the number of pixels occupied by small targets is very small, so the information of small targets is also very small, so that it is difficult for the detection layer to accurately detect small targets. And it is greatly reduced for the detection accuracy over long-distance traffic signs. We add 104×104 detection layer in multi-scale detection to increase the information of small target traffic signs, so as to improve the detection accuracy and speed. The

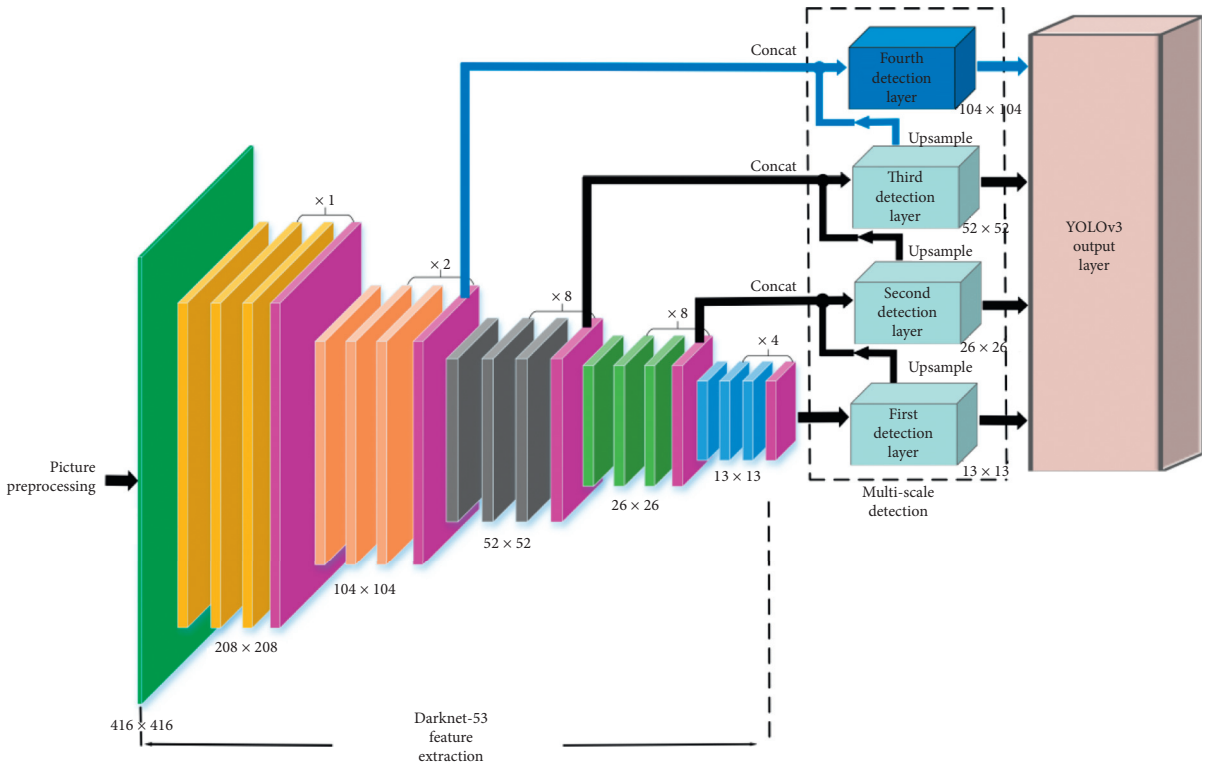

Figure 1: YOLOv3 structure diagram.

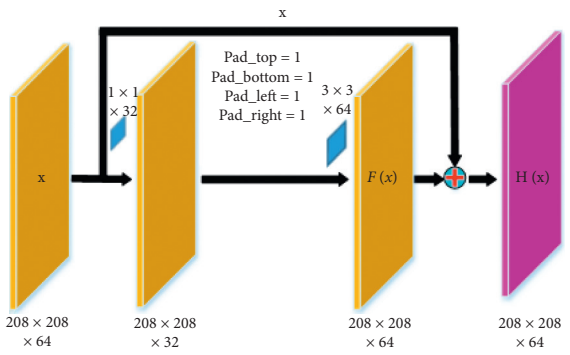

Figure 2: Residual structural diagram in Darknet-53.

104×104 detection layer is obtained by concatenating the up-sampled 52×52 detection layer with the 52×52 feature map in Darknet-53. This approach allows the new 104×104 detection layer to have more semantic information and local detail information to improve the detection accuracy of the target. And the 104×104 detection layer can generate a larger bounding box, which is helpful to detect small targets. It is expressed as (4) for the number of pixels of small target on a 104×104 feature graph.

$$N_{104} \leq \left(64 \times \frac{1}{9}\right) \times \left(64 \times \frac{1}{9}\right) < 50. \tag{4}$$

Obviously, the detection layer of 104×104 can detect more pixels, which improves the detection accuracy of the small target and helps to reduce the missed detection rate and false detection rate of small targets at long distance.

As shown in Figure 3, the fourth detection layer, an additional 104×104 large scale, is added to the multi-scale detection of YOLOv3. When the detection layer detects the target layer with the scale of 104×104, a 104×104 feature map is generated, and then the feature map is divided into 104×104 grid cells, in which the small targets are separated by grid cells. The advantage of the method is that it can detect and locate small targets more accurately. The 104×104 detection layer can also generate more predicted bounding boxes to make more accurate predictions of targets. This can avoid the missed detection of small targets. Then, we calculate the iou value of each predicted bounding box with ground truth

$$\text{iou}\left(a_m, b_i\right) = \frac{a_m \cap b_i}{a_m \cup b_i}, \tag{5}$$

where $\text{iou}(a_m, b_i)$ is the intersection ratio, $a_m \cap b_i$ is the overlap area between a_m and b_i, and $a_m \cup b_i$ is the total area of a_m and b_i. Finally, YOLOv3 removes redundant frames based on NMS and selects the best bounding box.

3.2. Soft-NMS. Traditionally, NMS only retains the best-performing bounding box and removes redundant bounding boxes. The scores of the redundant bounding boxes are forced to be settled as 0. The calculation method of the non-maximum suppression method in YOLOv3 is

$$S_i = \begin{cases} S_i, & \text{iou}\left(a_m, b_i\right) < N_t, \\ 0, & \text{iou}\left(a_m, b_i\right) \geq N_t, \end{cases} \tag{6}$$

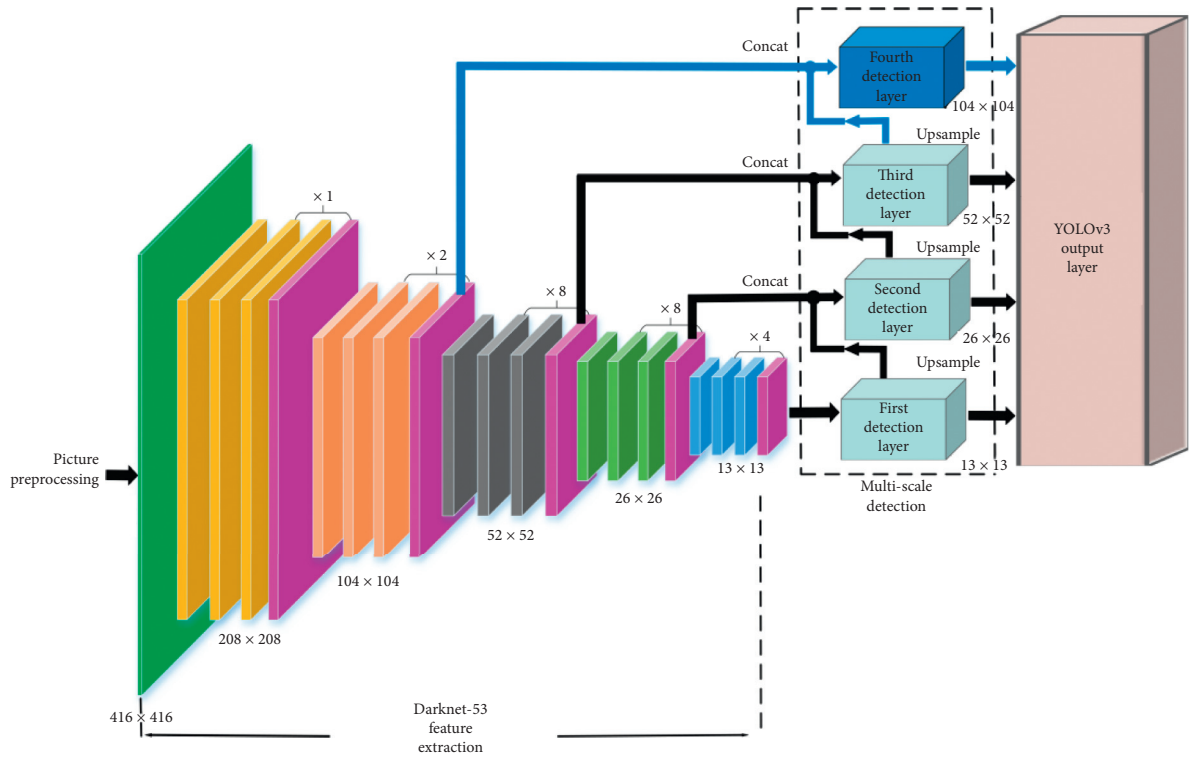

Figure 3: Framework of improved YOLOv3.

where S_i is the score of the i bounding box, a_m is the bounding box with the highest score, b_i is the currently unselected bounding box, $\text{iou}(a_m, b_i)$ is the intersection-parallel ratio between a_m and b_i, and N_t is the threshold.

The disadvantage of this method is that when two similar objects approach each other, the network only retains the boundary box of the objects with higher scores, but ignores the boundary box with lower scores. Therefore, the network output results are likely to regard these two objectives as one object.

As shown in Figure 4, two traffic signs photographed may overlap at a road corner. This will cause the boundary boxes of the two targets to overlap seriously. When the iou of two bounding boxes exceeds 0.5, the score of the bounding box with the lower score will be forced to be 0 by the NMS. This leads to missed targets with lower scores. In intelligent driving, the number of traffic signs on the road section prone traffic accidents is relatively large, and some traffic signs can cover other traffic signs sometimes. In this case, many phenomena such as the traffic signs in Figure 4 will be blocked and the driver cannot judge the traffic sign accurately, which will bring a lot of psychological pressure on the driver and thus increase the probability of a traffic accident. So, we replace NMS by soft-NMS in YOLOv3 to improve the recognition rate of overlapped targets.

Soft-NMS uses the linear weighting method to make the currently unselected bounding box get a lower score, instead of being directly forced to 0. This method avoids missing detection of the same kind of objects close to each other to a certain extent. We use Figure 5 to illustrate the specific calculation process of soft NMS, and the procedure is as follows. The first step is to sort all bounding boxes, and retain the one with the highest score. The second step is to calculate the iou between the remaining bounding box and the reserved bounding box. Third, we set a threshold of 0.5 and set the score of the remaining bounding boxes with iou greater than 0.5 to be $S_i(1 - \text{iou}(a_m, b_i))$. The fourth step repeats previous steps for all the bounding boxes until the scores of all the bounding boxes are updated. The specific expression of linear weighting can be written as follows:

$$S_i = \begin{cases} S_i, & \text{iou}(a_m, b_i) < N_t, \\ S_i(1 - \text{iou}(a_m, b_i)), & \text{iou}(a_m, b_i) \geq N_t. \end{cases} \tag{7}$$

By (7), the remaining bounding box with iou greater than 0.5 will get a lower score, and the score is $S_i(1 - \text{iou}(a_m, b_i))$ instead of 0. Compared with NMS, the suppression degree of the score of the remaining bounding box whose iou is greater than 0.5 in the soft-NMS becomes smaller. Figure 4(a) shows the result obtained through NMS. It can be seen that only one of the two targets is correctly detected, and the other target is missing. This is because the overlap between the second target and the first target is too large. The result of using soft-NMS is shown in Figure 4(b), which indicates that soft-NMS can correctly detect two targets.

4. Experiment and Results

4.1. Dataset. Our dataset comes from CSUST Chinese Traffic Sign Detection Benchmark (CCTSDB) [25, 26]. These images include streets traffic scenes, high-speed

(a) (b)

Figure 4: (a) Result picture detected by NMS. (b) Result picture detected by soft-NMS.

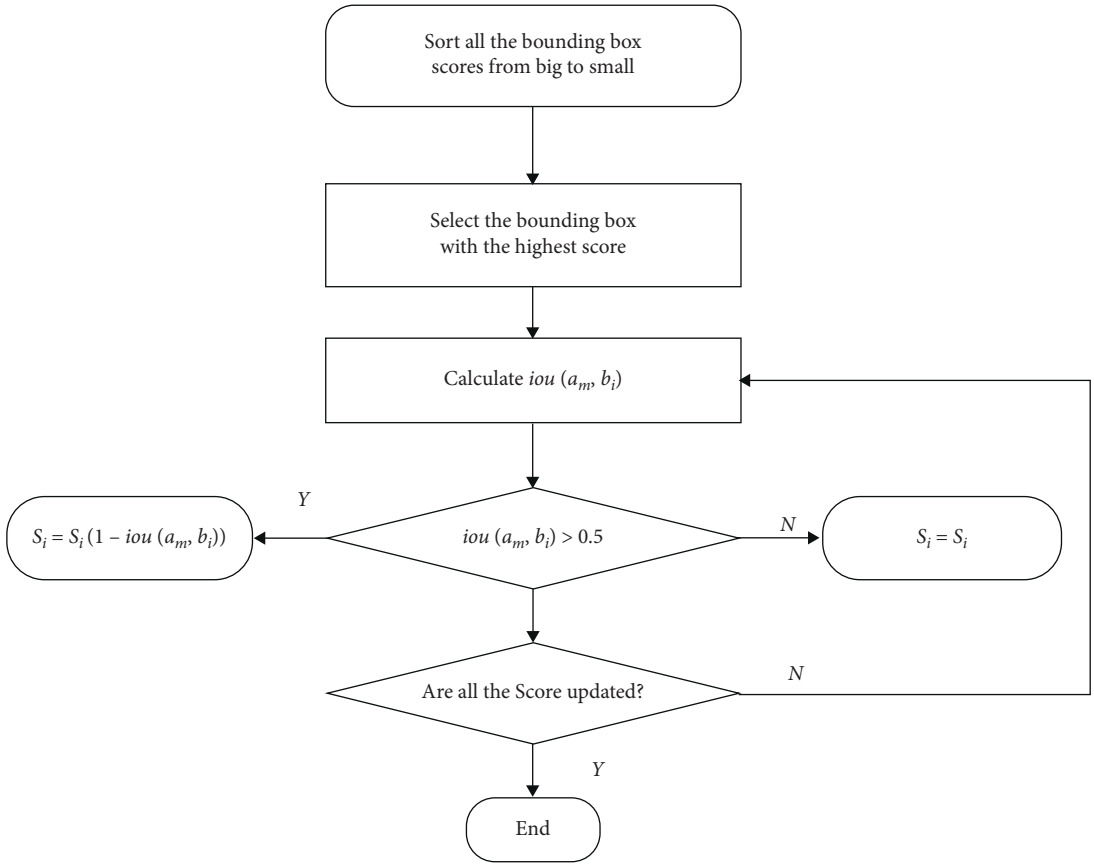

Figure 5: Soft-NMS calculation flowchart.

traffic scenes, rainy weather traffic scenes, evening traffic scenes, and backlit traffic scenes. However, CCTSDB only contains three categories of signs: prohibitory, mandatory, and warning. To further subdivide traffic signs, we randomly sampled 6300 images from CCTSDB and then relabel and reclassify the original three categories into ten new categories: prohibitory, mandatory, warning, speed limit, height limit, turn indicator, no sirens, no U-turns, lanes merge, and no parking. By this procedure, the clustered images contain various traffic scenes, which are helpful to improve the robustness of training. We relabeled these images with labelImg. Finally, we formed seven new categories to help improve the safety of assisted driving. And the final dataset contained 5000 training images and 1300 test images. These images are obtained from CCTSDB and are of different sizes, such as 1000×300, 1024×768, and 1280×720. Moreover, some images had been converted into 513×999 and 641×936 [24]. However, all the images will be resized to 416×416 after entering YOLOv3. As shown in Figure 6, the categories of the new dataset are prohibitory, mandatory, warning, speed limit, no parking, height limit, no sirens, no U-turns, lanes merge, and turn indicator.

In addition, this dataset contains many small-shaped traffic signs, and the traffic scene is more complicated. In Figure 7, the size of the original image is 1280×720, but the pixels of the traffic sign in the figure are 25×23 and 26×26 (smaller than 32×32) which shows that this dataset is suitable for training small targets.

Figure 6: Chinese traffic sign classification map.

Figure 7: Some samples in our dataset.

4.2. Experimental Process. The experimental environment is as follows: deep learning open source framework Darknet, Operating system Ubuntu 16.04, Machine learning system TensorFlow 1.3, CPU Intel Xeon E5 2678 V3 processor (×2), GPU GTX 1080 with 8 GB of memory (×4).

The test result is obtained by training the YOLOv3 model with a learning rate of 0.001 and a momentum of 0.09. The number of epochs in this experiment is 640. When training the network, we change the size of the training picture randomly. Compared with the single scale network training, multi-scale training is helpful for the network to predict the detection at different resolutions [12]. And we use K-Means clustering to re-cluster the training set to further improve the learning speed. In the improved YOLOv3, 12 anchor boxes are selected, and the specific parameters are as follows: (6, 11), (8, 20), (12, 19), (14, 30), (20, 34), (26, 39), (32, 53), (44, 60), (52, 178), (72, 98), (97, 129), and (142, 182). Finally, the experiment calculates the final weights over 50,200 iterations. Figure 8 shows the loss curves of the three networks in the iterative process. As can be seen from Figure 8(a), the final loss value of YOLOv2 fluctuates around 1.4. We can also see from Figures 8(b) and 8(c) that the final loss values of YOLOv3 and the improved YOLOv3 both fluctuate around 0.1. However, it is obvious that the floating range of the loss value in Figure 8(c) is lower than that in Figure 8(b). This indicates that the improved loss of YOLOv3 is more stable than the loss of YOLOv3 during training. This also means that the improved YOLOv3 has better detection effect on small targets than YOLOv3. Some specific experimental parameters are shown in Table 1.

4.3. Experimental Results. The improved YOLOv3 is trained and tested on our new dataset. At the same time, the training and testing of HOG + SVM, Faster R-CNN, YOLOv2, YOLOv3, YOLOv3 + fourth detection layer, and YOLOv3 + fourth detection layer + soft-NMS are run in the same environment In order to further verify that the improved YOLOv3 helps to detect small targets, we also selected 21 categories of targets with a large number of pictures in TT100 to conduct experiments [27]. TT 100 k is composed of 10,000 images, containing 100 classes of traffic signs with a resolution of 2048×2048 [28]. Some experimental results are shown in Figure 9.

The size of all images in Figure 9 is 847×635, and the minimum sizes of the target to be detected in the four pictures are 57×50, 43×39, 18×18, and 21×20. Figure 9(a) shows that when HOG + SVM is used to detect the six traffic signs in the four pictures, no traffic signs can be detected correctly. And the number of erroneously detected targets is 65. Faster R-CNN detected four traffic signs correctly. It missed 1 traffic sign, and it detected two traffic signs incorrectly (Figure 9(b)). In addition, we used YOLOV1, YOLOV1-tiny, YOLOv2, and YOLOv3 to detect four images. The experimental results show that YOLOV1 did not detect any sign correctly. YOLOV1-tiny

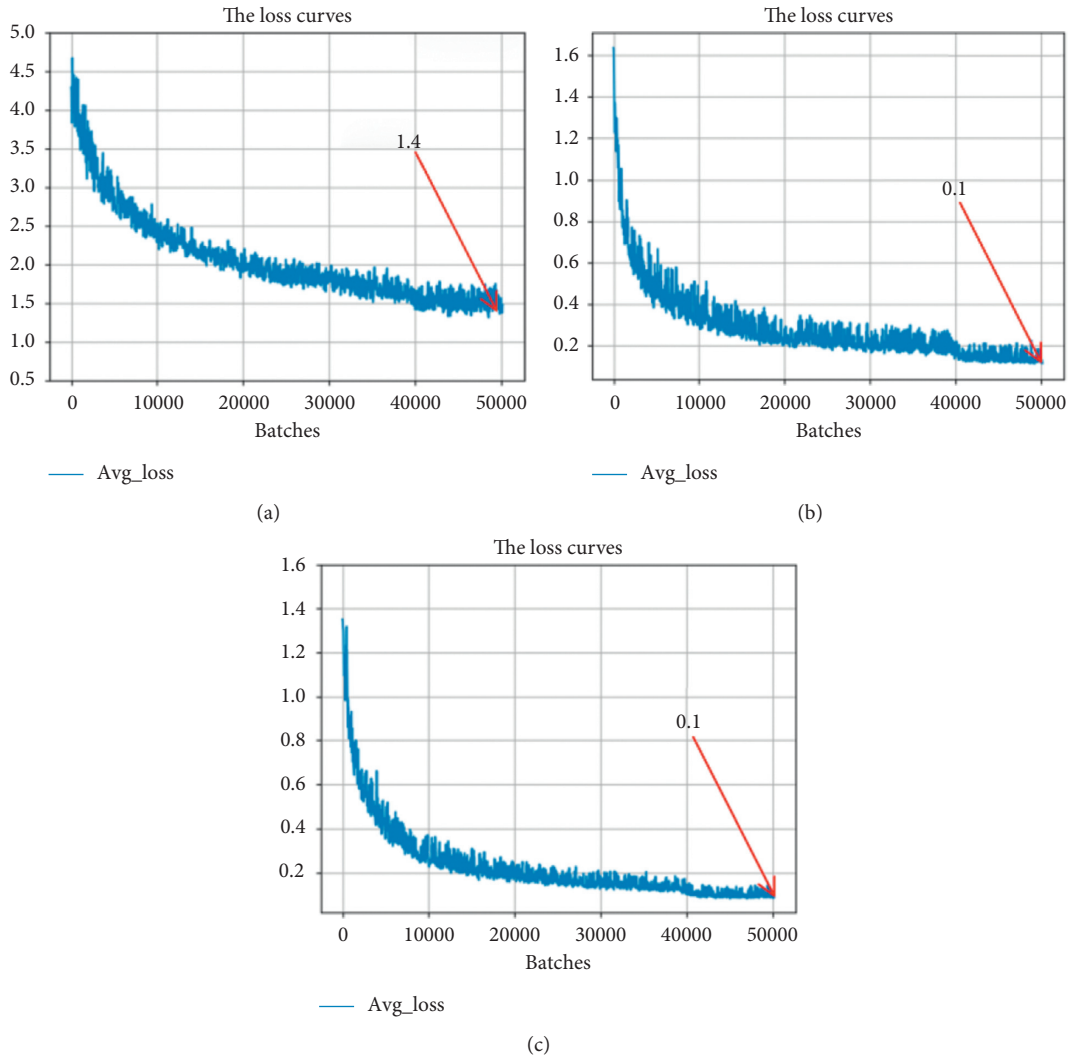

Figure 8: (a) Loss change process curve of YOLOv2. (b) Loss change process curve of YOLOv3. (c) Loss change process curve of improved YOLOv3.

Table 1: Comparison table.

Parameter name	Batch	Subdivisions	Width	Height	Momentum	Decay	Learning rate	Saturation	Exposure	Max_batches
Parameter value	64	16	416	416	0.9	0.0005	0.001	1.5	1.5	50200

only detected one sign correctly. YOLOv2 can only correctly detect two traffic signs while two signs are missed and incorrectly detected. YOLOv3 can detect three traffic signs correctly, and the numbers of missed and incorrectly detected traffic signs are one and two, respectively. Finally, we used the improved detection method based on YOLOv3 to detect traffic signs. We found that this method can detect six traffic signs in four pictures correctly. Table 2 shows that the mAP of the improved YOLOv3 on our test set is 98.8%, which is 5.7% higher than that of YOLOv3. Moreover, the FPS of the improved YOLOv3 can reach 43.8. This shows that the improved YOLOv3 can detect small traffic signs quickly and accurately for long distances during intelligent driving. However, our method will reduce the detection accuracy in darker environments. This may be solved by methods such as image enhancement. Of course, in addition to traffic signs, the obstacles in the environment must also be taken into consideration during the operation of the vehicle [29]. We will also study these issues in the future.

Figure 9: (a) Experimental test pictures of HOG + SVM. (b) Experimental test pictures of Faster R-CNN. (c) Experimental test pictures of YOLOv1. (d) Experimental test pictures of YOLOv1-tiny. (e) Experimental test pictures of YOLOv2. (f) Experimental test pictures of YOLOv3. (g) Experimental test pictures of improved YOLOv3.

Table 2: Comparison table.

Names	TT 100 k		Our dataset	
	mAP	FPS	mAP (%)	FPS
HOG + SVM [10]	0.1	2.2	4.9	12.4
Faster R-CNN [11]	50.3	3.6	74.9	6.20
YOLOv1 [13]	16.9	41.0	57.8	72.0
YOLOv1-tiny	11.5	54.9	38.6	168.0
YOLOv2 [14]	19.8	31.2	65.1	93.7
YOLOv3 [15]	52.9	25.6	93.1	49.2
YOLOv3 + fourth detection layer	54.2	13.3	98.6	43.9
Improved YOLOv3 (YOLOv3 + fourth detection layer + soft-NMS)	54.3	13.3	98.8	43.8

5. Conclusions

This paper detected Chinese traffic signs with the improved YOLOv3. We used four detection layers and soft-NMS to increase the accuracy of small target detection. After repeated training of the model through the training set, the improved YOLOv3 has a good detection effect in small traffic sign detection. Compared with the traditional detection method, the method we proposed has high detection speed and high detection precision, meeting the requirements of rapid and accurate detection in intelligent driving. However, compared to the ideal detection requirements, our network is prone to the missed detection of some vague targets. This issue will serve as our main research direction in the future.

Data Availability

The data used to support the findings of this study are currently under embargo while the research findings are commercialized. Requests for data, 6 months after publication of this article, will be considered by the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported in part by 2020 Anhui Provincial Engineering Laboratory on Information Fusion and Control of Intelligent Robot Open Project (no. ifcir2020002) and in part by the National Natural Science Foundation of China (nos. 61871003 and 11971034).

References

[1] C. Xiong, C. Wang, W. Ma, and Y. Shan, "A traffic sign detection algorithm based on deep convolutional neural network," in *Proceedings of the IEEE International Conference on Signal and Image Processing (ICSIP)*, pp. 676–679, Beijing, China, July 2016.

[2] Y. Tong and H. Yang, "Real-Time traffic sign detection method based on improved convolution neural network," *Laser & Optoelectronics Progress*, vol. 56, no. 7, pp. 123–129, 2019, in Chinese.

[3] X. Dai, X. Yuan, G. Le, and L. Zhang, "Detection method of traffic signs based on color pair and MSER in the complex environment," *Journal of Beijing Jiaotong University*, vol. 42, pp. 107–115, 2018, in Chinese.

[4] P. Cui and R. Zhang, "Research on road traffic sign recognition technology based on color-geometry modeling," *Internet of Things technology*, vol. 7, pp. 17-18, 2017, in Chinese.

[5] J. Ma, N. Yang, and X. Zhang, "Small target detection in foggy image combined with low-rank and structrued sparse," *Computer Engineering and Applications*, vol. 54, pp. 176–182, 2018, in Chinese.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *Institute of Electrical and Electronics Engineers Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[7] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, Columbus, OH, USA, June 2014.

[8] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, Chile, June 2015.

[9] S. Ren, K. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *Institute of Electrical and Electronics Engineers Transactions Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[10] C. Yao, F. Wu, H. Chen, X. Hao, and Y. Shen, "Traffic sign recognition using HOG-SVM and grid search," in *Proceedings of the 2014 12th International Conference on Signal Processing (ICSP)*, pp. 962–965, HangZhou, China, October 2014.

[11] Z. Zuo, K. Yu, Q. Zhou, X. Wang, and T. Li, "Traffic signs detection based on faster R-CNN," . in *Proceedings of the 2017 IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW)*, pp. 286–288, Atlanta, GA, USA, June 2017.

[12] S. Song, Z. Que, J. Hou, S. Du, and Y. Song, "An efficient convolutional neural network for small traffic sign detection," *Journal of Systems Architecture*, vol. 97, pp. 269–277, 2019.

[13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, Las Vegas, NV, USA, June 2016.

[14] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6517–6525, Honolulu, HI, USA, July 2017.

[15] J. Redmon and A. Farhadi, *YOLOv3: An Incremental Improvement*, 2018.

[16] J. Tao, H. Wang, X. Zhang, X. Li, and H. Yang, "An object detection system based on YOLO in traffic scene," in *Proceedings of the 2017 6th International Conference on Computer Science and Network Technology (ICCSNT).*, pp. 315–319, Dalian, China, October 2017.

[17] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297, Berkeley, CA, USA, January 1967.

[18] P. Garg, D. R. Chowdhury, and V. N. More, "Traffic sign recognition and classification using YOLOv2, faster RCNN and SSD," in *Proceedings of the 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1–5, Kanpur, India, January 2019.

[19] J. Khan, Y. Chen, Y. Rehman, and H. Shin, "Performance enhancement techniques for traffic sign recognition using a deep neural network," *Multimedia Tools and Applications*, vol. 79, no. 29-30, pp. 20545–20560, 2020.

[20] A. Neubeck and L. Van Gool, "Efficient non-maximum suppression," in *Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)*, pp. 850–855, Hong Kong, China, August 2006.

[21] N. Bodla, B. Singh, R. Chellappa, and L. S. Davis, "Soft-NMS improving object detection with one line of code," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 5561–5569, Venice, Italy, October 2017.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition Las Vegas*, Las Vegas, NV, USA, June 2016.

[23] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *Institute of Electrical and Electronics Engineers Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994.

[24] Z. Zhang, H. Wang, J. Zhang, and Y. Wei, "A vehicle real-time detection algorithm based on YOLOv2 framework," *Real-time Image & Video Processing*, vol. 2018, 2018.

[25] J. Zhang, X. Du, and J. Xin, "Spatial and semantic convolutional features for robust visual object tracking," *Multimedia Tools and Applications*, vol. 77, pp. 1–21, 2018.

[26] J. Zhang, Q. Huang, H. Wu, and Y. Liu, "Effective traffic signs recognition via kernel PCA network," *International Journal of Embedded Systems*, vol. 10, no. 2, pp. 120–125, 2018.

[27] Z. Zhu, L. Dun, S. Zhang et al., "Traffic-sign detection and classification in the wild," in *Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition*, IEEE, Seattle, WA, USA, June 2016.

[28] Z. Meng, X. Fan, X. Chen et al., "Detecting small signs from large images," in *Proceedings of the 2017 IEEE International Conference on Information Reuse and Integration*, pp. 217–224, San Diego, CA, USA, August 2017.

[29] G. Qi, H. Wang, M. Haner, C. Weng, S. Chen, and Z. Zhu, "Convolutional neural network based detection and judgement of environmental obstacle in vehicle operation," *CAAI Transactions on Intelligence Technology*, vol. 4, pp. 80–91, 2019.