Journal of Computers Vol. 33 No. 6, December 2022, pp. 181-192
doi: 10.53106/199115992022123306015

YOLOv4-A: Research on Traffic Sign Detection Based on Hybrid Attention Mechanism

Songlin Yin[1], Fei Tan[2*]

[1] School of Automation and Information Engineering, Sichuan University of Science & Engineering, Zigong 643000, China
[2]Artificial Intelligence Key Laboratory of Sichuan Province, Yibin 644000, China
13951423843@139.com, 549152121@qq.com

Received 12 March 2022; Revised 7 June 2022; Accepted 6 July 2022

Abstract. Aiming at the problem of false detection and missed detection in the traffic sign detection task, an improved YOLOv4 detection algorithm is proposed. Based on the YOLOv4 algorithm, the Efficient Channel Attention Module (ECA) and the Convolutional Block Attention Module (CBAM) are added to form YOLOv4-A algorithm. At the same time, the global K-means clustering algorithm is used to regenerate smaller anchors, which makes the network converge faster and reduces the error rate. The YOLOv4-A algorithm re-calibrates the detection branch features in the two dimensions of channel and space, so that the network can focus and enhance the effective features, and suppress the interference features, which improves the detection ability of the algorithm. Experiments on the TT100K traffic sign dataset show that the proposed algorithm has a particularly significant improvement in the performance of small target detection. Compared with the YOLOv4 algorithm, the precision and mAP@0.5 of the proposed algorithm are increased by 5.38% and 5.75%.

Keywords: traffic sign detection, YOLOv4, K-means, attention mechanism

1 Introduction

Traffic sign detection can not only provide effective road condition data support for assisted driving systems, but also avoid cumbersome and error-prone manual annotations in building high-precision maps. Therefore, in-depth research on the traffic sign detection system not only has great practical value in improving road safety, but also can promote the development of driverless technology. In recent years, with the development of convolution neural network in the field of computer vision, traffic sign detection algorithm based on deep learning has also made great progress. Existing detection methods can be divided into two-stage methods and one-stage methods. The two-stage method represented by Faster R-CNN [1] uses RPN to generate suggestion boxes at the feature level by sharing the convolution features, then uses the convolution features of the suggestion box area to classify and locate the target boxes, it has the characteristics of high accuracy but low speed. The one-stage target detection method represented by YOLO [2] and SSD [3] unifies the positioning and recognition tasks of the target frame according to the logic of regression, and is completed by the convolutional neural network in the output layer, it has the characteristics of high speed but low precision.

At present, many experts and scholars have made good achievements in traffic sign detection. Rajendran et al. [4] based on RetinaNet [5], used ResNet [6] with deeper layers as the basic network, and used the deconvolution module at the bottom of the network to enrich the semantic information of features, and finally obtained 96.7% on the GTSDB traffic dataset, but this method will introduce a large number of additional parameters. Yang et al. [7] used a multi-scale fully convolutional network DMS-Net to detect traffic signs of different scales and introduced an online difficult sample mining strategy, finally achieved 99.88% accuracy and 96.61% recall on the STSD dataset. Meng et al. [8] divided each image into small images of 200 pixels × 200 pixels on the basis of the image pyramid, sent them to the SSD network for target detection and trained an SOS network that was sensitive to small targets. However, image pyramid and sub-image division operations are also required during testing, which reduces the real-time performance of the algorithm. The above work improves the performance of traffic sign detection algorithms from different perspectives.

In the actual traffic sign detection scene, the image background is complex and diverse, there are various billboards, interfering objects and other prompt signs. These fake traffic signs are easily confused with the real traffic signs in shape and color, which can easily lead to false detection [9]. In addition, in order to obtain road information in advance, the traffic signs captured by the vehicle-mounted camera generally have a small absolute pixel

* Corresponding Author

size, and the relative proportion of the entire image is also very small, and it contains less effective information and more noise. Therefore, in the case of ambiguity, it is easy to miss detection [10]. In view of the above problems, this paper uses the YOLOv4 algorithm as the basis, adds the ECA module in the downsampling stage of the PAN structure, and adds the CBAM module before the three prediction heads to form the YOLOv4-A algorithm to reduce false detection and missed detection in the task of traffic sign detection. At the same time, the global K-means clustering algorithm is used to regenerate smaller anchors to speed up the convergence of the network. The model trained by the final algorithm achieved good results on the TT100K test set.

2 YOLOv4 Algorithm

The YOLOv4 algorithm belongs to a single-stage target detection algorithm, which integrates target classification and target localization in the same network for prediction, realizing “end-to-end” detection [11]. The network structure of the YOLOv4 algorithm is shown in Fig. 1. Its backbone feature extraction network adopts CSPDarknet53 [12], CSPDarknet53 is improved on the basis of Darknet53 [8]. It uses CSPNet [13] in the Resblock_body structure, which can further enhance the learning ability of the network, on the premise of ensuring accuracy reduce the memory cost and computational bottleneck, and also use the Mish [14] activation function to replace the original LeakyReLU activation function. The Mish activation function is smoother than LeakyReLU in the negative region, which is conducive to the calculation and update of the gradient, and can obtain better accuracy and generalization ability. Spatial pyramid pooling structure introduced after trunk feature extraction network, and the SPP [15] structure is processed by combining different maximum pooling layers, it can increase the receptive field to separate out significant contextual features. The path aggregation network adds a bottom-up path enhancement structure on the basis of the feature pyramid network, shortens the information path from low-level features to high-level features, and speeds up the flow of low-level information. The detection head follows the head in YOLOv3, which consists of 3 × 3 and 1 × 1 convolutional layers, and predicts the result of the feature layer output after PANet processing [16].

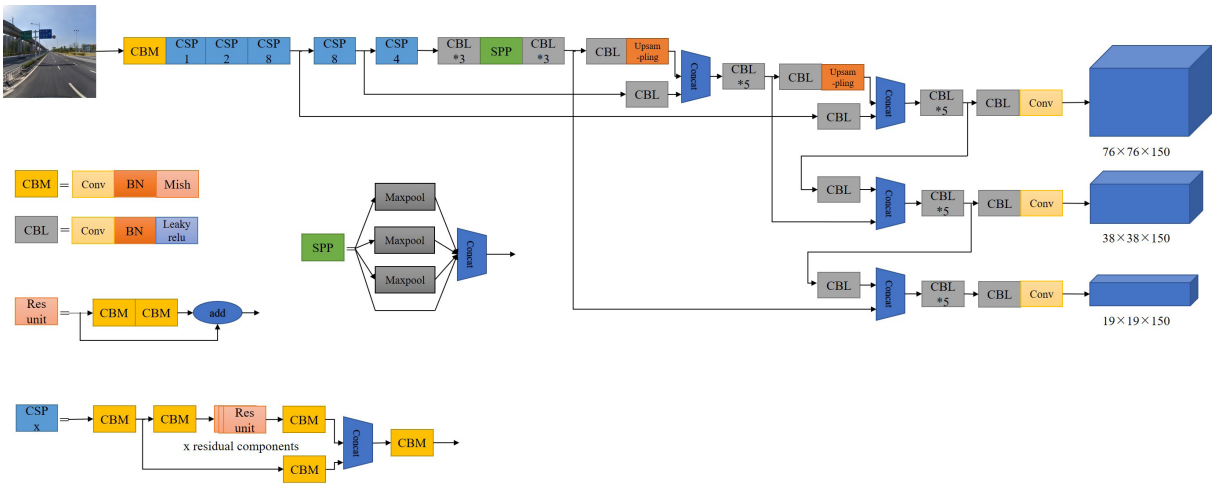

Fig. 1. YOLOv4 network structure

3 YOLOv4-A Traffic Sign Detection Network

3.1 Anchor Prediction Based on Global K-means Clustering

In the anchor-based target detection algorithm, the anchor is set manually or with prior training knowledge in most cases. However, the size of the artificially set anchor is usually difficult to completely match the data, resulting in suboptimal training results. For example, in RCNN, SSD, Faster-RCNN, 9 anchors with different aspect ratios and sizes are usually set [17]. The manually designed anchor size is generally not suitable for the training of the dataset. In most anchor-based target detection algorithms, the anchor size set before training is usually obtained by K-means clustering. The accuracy of the results of the K-means algorithm is affected by the initial clustering selection, which will lead to unstable results of the anchors, resulting in the final anchors not being the

optimal size [18]. The steps of the K-means algorithm are shown in Fig. 2.

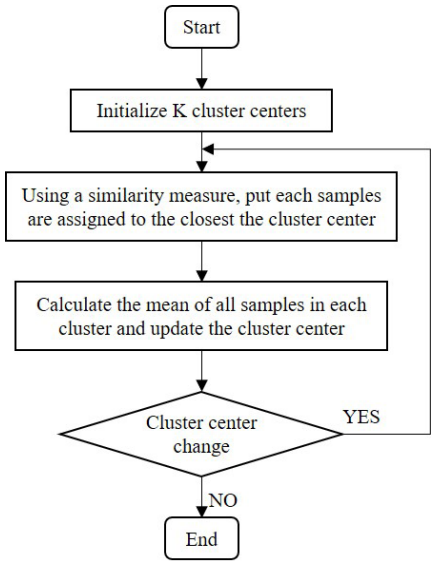

Fig. 2. K-means clustering steps

The anchor provided by the original YOLOv4 algorithm is obtained by clustering the COCO dataset, and the COCO dataset picture is very different from the target size and target type in the traffic sign dataset picture. The value of the anchor will affect the performance of the network and an appropriate anchor size can make the network converge faster and reduce the error rate. Therefore, this paper uses the global K-means clustering algorithm to regenerate the anchor, and the flowchart is shown in Fig. 3.

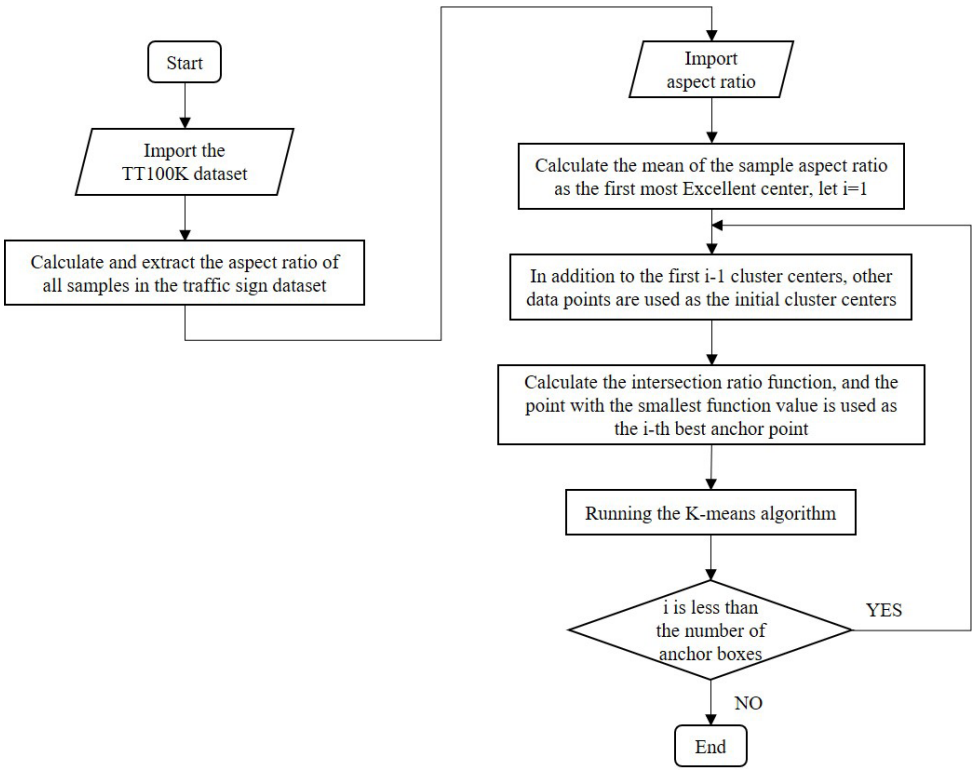

Fig. 3. Initial anchor prediction process by global K-means clustering

3.2 Attention Mechanism

The attention mechanism has been proved to be effective in improving network performance in many computer vision tasks. This method simulates the process of human brain extracting external information, that is, the human visual system will generate local focus on certain areas on the image, and by focusing on the focal areas to invest more attention and get effective details [19].

A. Convolutional Block Attention Module

The convolutional attention module is a simple and effective attention module for feed-forward convolutional neural networks, the principle is shown in Fig. 4. For the feature map extracted from the feature, the CBAM infers the attention map in turn along the two independent dimensions of channel and space, and then multiplies the attention map with the input feature map for adaptive feature optimization. The formula is shown in (1) (2). Since CBAM is a lightweight general-purpose module, it can be seamlessly integrated into any convolutional network structure with negligible computational effort, and can be trained end-to-end together with the underlying convolutional network [20].

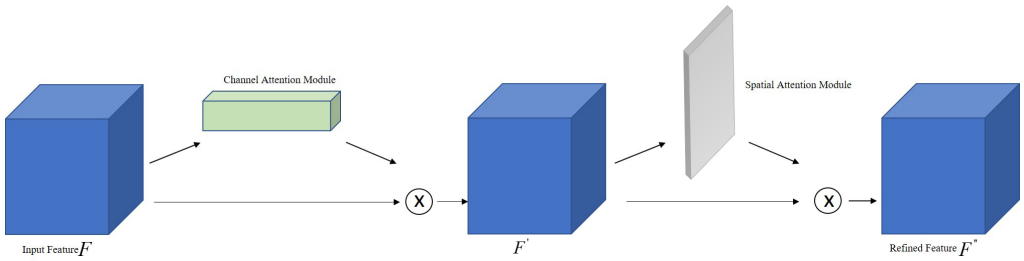

Fig. 4. Convolutional attention module

$$F' = F \times M_c(F), \tag{1}$$

$$F'' = F' \times M_s(F'), \tag{2}$$

The channel attention module is shown in Fig. 5. The obtained feature maps are sent to the fully connected layer after maximum pooling and average pooling respectively. Finally, the obtained results are added and activated through the activation function, and the channel attention's value is obtained. The channel attention mechanism is to compress the feature map in the spatial dimension to obtain a one-dimensional vector and then operate. Average pooling and max pooling can be used to aggregate spatial information of feature maps, compress the spatial dimension of input feature maps, and element-wise sum and merge to produce channel attention maps. As far as a picture is concerned, channel attention focuses on what content on this picture is important. Average pooling has feedback for every pixel on the feature map, while max pooling only has gradient feedback where the response is the largest in the feature map when performing gradient backpropagation calculations. The expression of the channel attention mechanism is shown in (3).

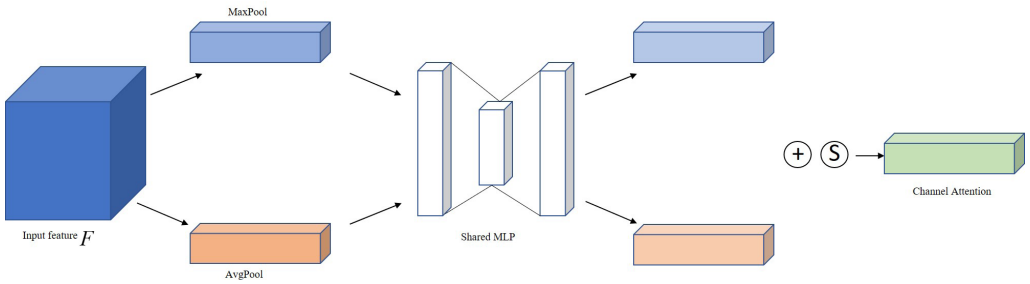

Fig. 5. Channel attention module

$$M_c(F)=\sigma\left(MLP\left(AvgPool(F)\right)+MLP\left(MaxPool(F)\right)\right), \tag{3}$$

The spatial attention module is shown in Fig. 6. The feature map output by the channel attention module is used as the input feature map of this module. First, do a channel-based max pooling and average pooling, and then do a splicing operation based on the channel, after a convolution operation, reduce the dimension to 1 channel, and finally get the spatial attention weight through the activation function. The spatial attention mechanism is to compress the channel, and the specific expression is shown in (4), where σ is the activation operation, and 7×7 represents the size of the convolution kernel.

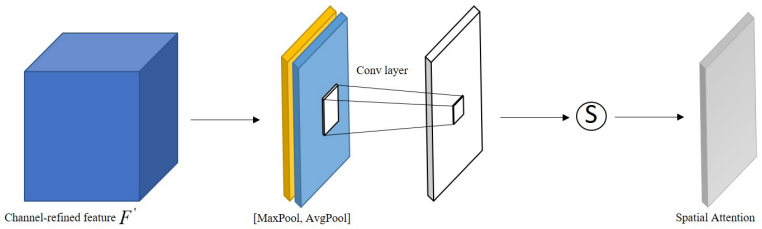

Fig. 6. Spatial attention module

$$M_s(F')=\sigma\left(f^{7\times7}\left(\left[AugPool\left(F'\right);MaxPool\left(F'\right)\right]\right)\right). \tag{4}$$

B. Efficient Channel Attention Module

The ECA module can be seen as an improved version of the SE [21] module, as shown in Fig. 7. The authors of ECA believe that SE brings side effects to the prediction of the channel attention mechanism, capturing all channel dependencies is inefficient and unnecessary, while convolution has good cross-channel information acquisition ability, so the ECA module uses 1D convolution replaces the two full connections of the SE module. The size of the convolution kernel of 1D convolution will affect the coverage of cross-channel interaction, so the choice of the size of the 1D convolution kernel k becomes very important. Although k can be adjusted manually, it will waste a lot of time and energy. k is nonlinearly proportional to C, the larger C is, the stronger the long-term interaction is, whereas the smaller C is, the stronger the short-term interaction is, as shown in equation (5).

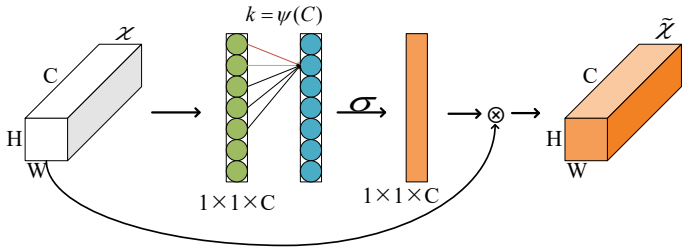

Fig. 7. ECA module structure diagram

$$C=\phi(k)=2^{(\gamma\times k-b)}, \tag{5}$$

When the channel dimension C is determined, the convolution kernel size k is calculated by formula (6):

$$k=\psi(C)=\left|\frac{\log_2(C)}{\gamma}+\frac{b}{\gamma}\right|_{odd}, \tag{6}$$

where γ, b are adjustment parameters, and $|t|_{odd}$ represents the nearest odd t.

3.3 YOLOv4-A Algorithm

The network proposed in this paper introduces the Efficient Channel Attention Module (ECA) and the Convolutional Block Attention Module (CBAM) into the detection branch of the YOLOv4 network to form the YOLOv4-A network, whose structure is shown in Fig. 8. The network first extracts features through the basic semantic feature network, and introduces the ECA module to re-calibrate the multi-scale features between channels when the PAN structure is down-sampled, so as to enhance the effective channel features and suppress redundant channel features. Then, the CBAM module is added in front of the three detection heads of the network, and the features are re-calibrated in the spatial dimension supervised to achieve the purpose of strengthening the features of the effective area, and suppressing the features of the interference area. Finally, object detection is performed on the obtained attention features.

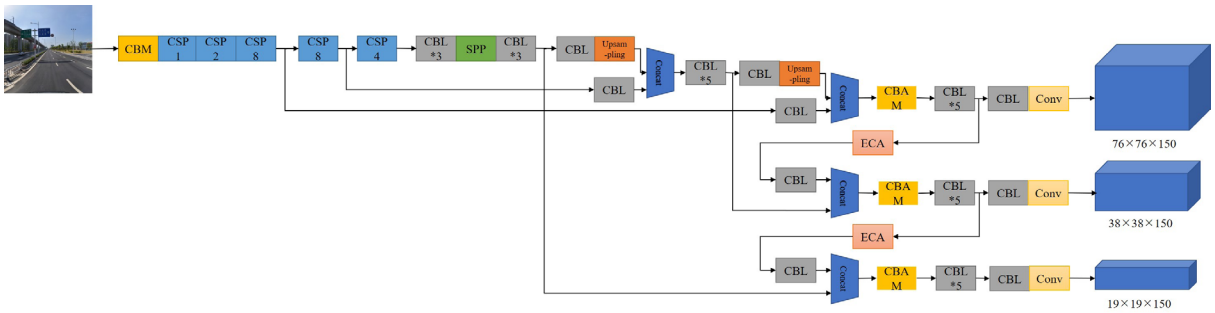

Fig. 8. YOLOv4-A network structure

4 Experiment and Result Analysis

4.1 Dataset

The experiments in this paper mainly use the TT100K [22] traffic sign dataset jointly produced by Tencent and Tsinghua University. The main reason for selecting this experimental set is that the data set contains a wide range of traffic signs and rich scenes. There are 221 types of traffic signs in the TT100K dataset, and the total number of targets is 26,349, both of which greatly exceed the datasets such as GTSDB, STS, and LISA. However, object detection on the TT100K dataset is challenging. For example, the small objects in this dataset have two problems: small absolute scale and small relative scale. The small absolute scale of the target means that the real scale of the traffic sign is small, that is, the pixel area occupied is small, which makes the obtained image target ambiguous, with less information and more noise, which makes the model detection difficult. The relative small scale of the target means that the image contains a lot of irrelevant background information, because the number of traffic signs in the image accounting for no more than 2% of the pixel area of the entire image exceeds 24,970, accounting for 94.7% of the total number of targets. Therefore, compared with other public datasets, the challenge of the TT100K dataset is more difficult. Part of the dataset in this experiment is shown in Fig. 9, which covers traffic signs of different sizes, angles, lighting conditions, and damage degrees, objectively reflecting the situation of traffic signs under realistic conditions.

Since there are many types in the TT100K dataset, the number of which is very small, and the distribution is extremely uneven, if 221 types of traffic signs are trained, the training effect will be unsatisfactory. Therefore, this paper selects 45 categories with more than 50 samples in the 1000 datasets through the python script for experiments, and some types are shown in Fig. 10. The distribution of the datasets is shown in Table 1.

Fig. 9. Pictures of some datasets

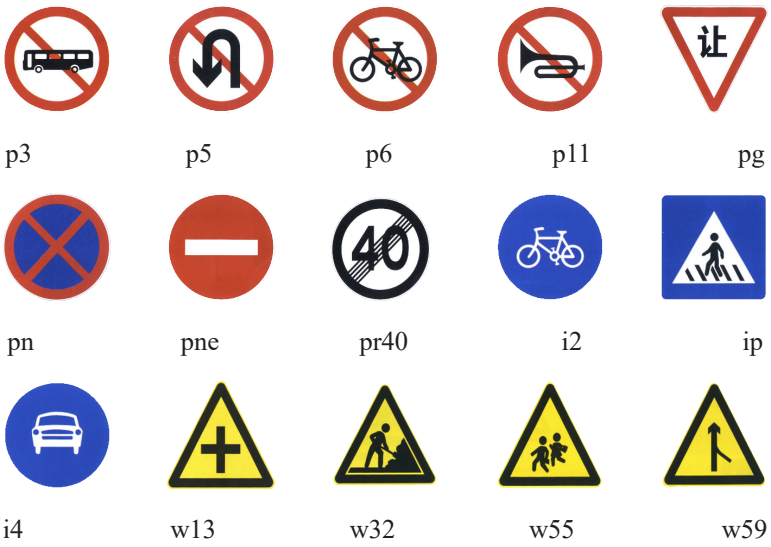

Fig. 10. Types of traffic signs in the experimental part

Table 1. Data set distribution

Total number of datasets	Training set	Valid set	Test set
10000	80%	10%	10%

4.2 Experimental Parameters and Evaluation Indicators

Experiments use a NVIDIA Geforce1080Ti graphics card with a memory size of 11GB to train and test all models. During the training process, due to the memory size limitation, set the batchsize to 8, the Decay to 0.0005, the Momentum to 0.9, the initial learning rate to 0.1, and the maximum number of training rounds is set to 200 epochs. A larger learning rate in the early stage of training can speed up the convergence, but an excessive learning rate in the later stage of training will cause the results to exceed the optimal value and cannot be fitted. Therefore, the current learning rate is attenuated by 10 times the previous learning rate every 50 epochs. The loss diagram of

the improved algorithm in this paper is shown in Fig. 11, when the training reaches 180 epochs, the loss tends to be stable and the model converges.

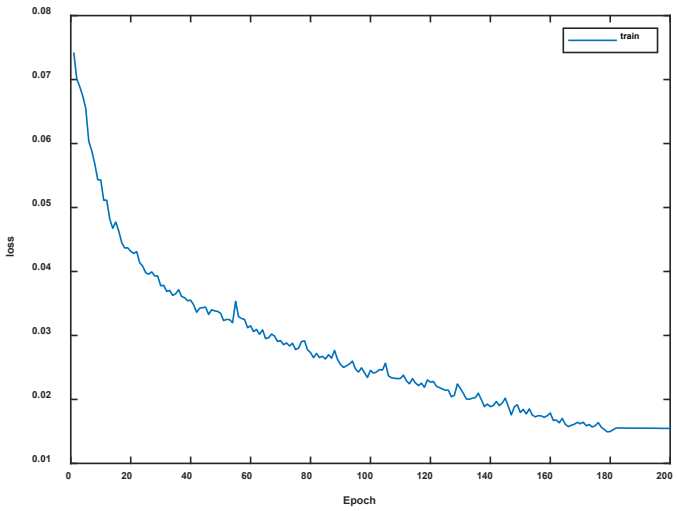

Fig. 11. The loss curve of the improved algorithm

The model evaluation index used in this paper is consistent with the method provided by Zhu [23], the publisher of the TT100K dataset, and a fixed IOU threshold and confidence threshold are used to judge whether the detection result is correct. Then calculate the precision, recall, mean Average Precision (mAP) and Frame Per Second (FPS) of the prediction results to measure the target classification ability and target detection ability of the model. Its calculation formula is as follows:

$$\Pr ecision = \frac{TP}{TP + FP}, \tag{7}$$

$$\mathrm{Re}\, call = \frac{TP}{TP + FN}, \tag{8}$$

$$mAP = \frac{1}{N}\sum AP_c, \tag{9}$$

When calculating the two indicators of Precision and Recall of the model, it is first necessary to divide the detection results into true positive examples (TP), true negative examples (TN), false positive examples (FP) and false negative examples (FN) according to the true labels 4 classes [24]. Precision means that the number of correctly detected samples accounts for the proportion of the total detected samples, which can reflect the classification ability of the model to the target. Recall means that the number of correctly detected samples accounts for the real samples, it can reflect the detection ability of the model to the target. mAP represents the mean of the average accuracy of all detection classes.

4.3 Analysis of Experimental Results

This paper conducts an intuitive evaluation of the confusion matrix to see the performance of the improved algorithm on the test set. The normalized confusion matrix is shown in Fig. 12. It can be seen from the matrix that the correct detection rate of il100 is the highest, while the correct detection rate of ph5 is lower. This is due to the fact that ph5 has too few samples in the data set, resulting in poor training effect, but in general, the improved algorithm in this paper is effective for traffic sign detection.

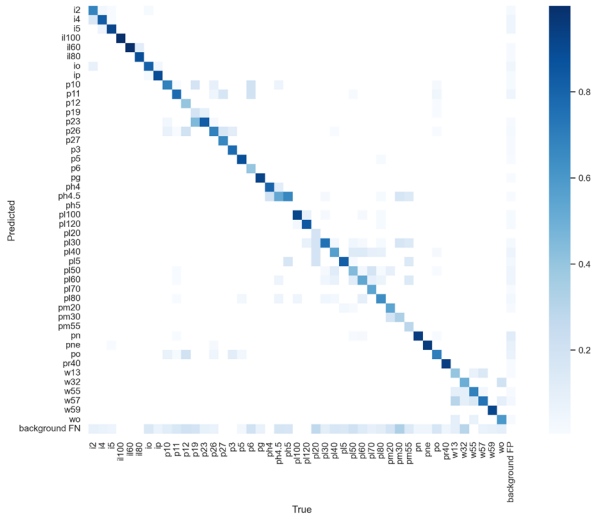

Fig. 12. Normalized confusion matrix

In order to demonstrate the effectiveness of the above improved method, this paper conducts ablation experiments on all scale targets of 45 categories in the TT100K dataset, and validates different modules by combining the aforementioned global K-means clustering a priori boxes and adding an attention mechanism to verify the improvement of network. The test results are shown in Table 2. The experimental results show that the global K-means clustering a priori box and the addition of the attention mechanism both improve the performance of the model, and improve the mAP by 0.32%, 0.74% and 0.47%. The improved algorithm after integrating all modules achieved the best results on P, R and mAP, which were improved by 5.38%, 5.84% and 5.75% compared with YOLOv4. It can be seen that the improved algorithm can significantly improve the network performance.

Table 2. Ablation experiment results on TT100K dataset

Models	Global K-means clustering prior box	Add CBAM module	Add ECA module	P (%)	R (%)	mAP@0.5 (%)
YOLOv4				70.22	69.71	70.57
A	√			71.45	70.09	70.89
B		√		72.51	72.16	71.31
C			√	71.65	71.50	71.04
D	√	√		72.65	72.52	72.04
E	√		√	73.22	73.15	72.78
F		√	√	74.52	74.50	74.85
Improve algorithm	√	√	√	75.60	75.55	76.32

Under the same experimental conditions, in order to verify the effectiveness of the algorithm, this paper chooses to compare with the classic algorithms R-CNN and Faster R-CNN in the two-stage target detection algorithm, and the classic algorithms SSD, YOLOv3 and YOLOv4-tiny in the single-stage target detection algorithm, and take the P, R, mAP and FPS as the evaluation criteria, and the comparative experimental results are shown in Table 3. It can be seen from Table 3 that the mAP of the algorithm in this paper is 8.08% higher than that of the second YOLOv4-tiny. In terms of FPS, although the algorithm in this paper is 29 images per second less than YOLOv4-tiny, the mAP is higher than it, and it also reaches the minimum requirements for real-time detection and is 4 images per second higher than YOLOv3. It can be seen that the improved algorithm performs well in both the accuracy and real-time performance of the network model.

Table 3. Comparison experiment of different algorithms

Models	P (%)	R (%)	mAP@0.5 (%)	FPS (f/s)
R-CNN	57.31	56.51	58.04	10
Faster R-CNN	66.45	65.15	67.85	18
SSD	60.29	60.75	61.46	25
YOLOv3	62.35	61.14	63.80	35
YOLOv4-tiny	67.15	66.74	68.24	68
Improve algorithm	75.60	75.55	76.32	39

Finally, in order to visually see the performance of the model trained by the improved algorithm in this paper, 3 images were randomly selected from the experimental data set in this paper to be tested with YOLOv4 and the improved algorithm respectively. The detection results are shown in Fig. 13.

YOLOv4 vs Ours

YOLOv4 vs Ours

YOLOv4 vs Ours

Fig. 13. Detection effect comparison pictures

Comparing Fig. 13(a), it can be seen that the YOLOv4 algorithm only detected 3 traffic signs in the figure and missed pl60, while the improved algorithm in this paper detected all 4 traffic signs correctly. In Fig. 13(b), there is only one traffic sign, the YOLOv4 algorithm mistakenly detects pl60 as pl40, and the improved algorithm avoids this problem. Finally, the results of Fig. 13(c) are more obvious. The improved algorithm in this paper identifies all p5 and p10 correctly and with high confidence, and the original algorithm not only falsely detected p10 as p5, but also falsely detected the logo on the billboard as pne. In summary, the prediction effect of the improved algorithm in this paper is significantly better than the original algorithm, so the improved algorithm is more capable of detecting traffic signs than YOLOv4.

5 Conclusion

This paper mainly introduces the improved YOLOv4 traffic sign detection algorithm. Aiming at the problem of false detection and missed detection in the traffic sign detection task, the ECA module is added in the downsampling stage of the PAN structure of the original algorithm, and the CBAM module is added before the three prediction heads, so that the network focuses its attention on useful information, which improves the detection ability of the algorithm. At the same time, the global K-means clustering algorithm is used to regenerate smaller anchors to speed up the network convergence. Then through ablation experiments, it is verified that the improved scheme proposed in this paper is improved compared with the original algorithm. Finally, the results of the improved model are compared with the five algorithms trained by R-CNN, Faster R-CNN, SSD, YOLOv3 and YOLOv4-tiny, the results show that the model trained by the improved algorithm is effective in traffic sign detection accuracy, false detection and missed detection, which has certain practical significance for improving driving safety and promoting the development of driverless technology. However, the detection accuracy of traffic signs in this paper is not very high, the next work is to study the problem that the traffic sign occupies a small position in the image and the detection accuracy is not high under the premise.

6 Acknowledgement

Thanks to the National Natural Science Foundation of China (61902268); Sichuan Provincial Science and Technology Program (2019YFSY0045) for funding this paper.

References

[1] S.-Q. Ren, K.-M. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, IEEE Transactions on Pattern Analysis and Machine Intelligence 39(6)(2017) 1137-1149.
[2] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: Proc. 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[3] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.-C. Berg, SSD: single shot multi box detector, in: Proc. 2016 European Conference on Computer Vision, 2016.
[4] S.-P. Rajendran, L. Shine, R. Pradeep, S. Vijayaraghavan, Fast and accurate traffic sign recognition for self-driving cars using Retina Net based detector, in: Proc. 2019 International Conference on Communication and Electronics Systems, 2019.
[5] T.-Y. Lin, P. Goyal, R. Girshick, K.-M. He, P. Dollar, Focal loss for dense object detection, in: Proc. 2017 IEEE International Conference on Computer Vision, 2017.
[6] K.-M. He, X.-Y. Zhang, S.-Q. Ren, J. Sun, Deep residual learning for image recognition, in: Proc. 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[7] Y.-C. Yang, S. Liu, W. Ma, Q.-Y. Wang, Z. Liu, Efficient traffic-sign recognition with scale-aware CNN, in: Proc. 2017 British Machine Vision Conference, 2017.
[8] Z.-B. Meng, X.-C. Fan, X. Chen, M. Chen, Y. Tong, Detecting small signs from large images, in: Proc. 2017 IEEE International Conference on Information Reuse and Integration, 2017.
[9] K.-M. He, G. Gkioxari, P. Dollar, R. Girshick, Mask R-CNN, IEEE Transactions on Pattern Analysis and Machine Intelligence 42(2)(2020) 386-397.
[10]T.-Y. Lin, P. Goyal, R. Girshick, K.-M. He, P. Dollar, Focal loss for dense object detection, in: Proc. 2017 IEEE International Conference on Computer Vision, 2017.
[11]R.-L. Gai, N. Chen, H. Yuan, A detection algorithm for cherry fruits based on the improved YOLO-v4 model, Neural Computing and Applications (2021) 1-12.
[12]C.-Y. Wang, H. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, I.-H. Yeh, CSPNet: a new backbone that can enhance learning capability of CNN, in: Proc. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops(CVPRW), 2020.

[13]J. Redmon, A. Farhadi, Yolov3:an incremental improvement, in: Proc. 2018 IEEE Conference on Computer Vision and Pattern Recognition, 2018.
[14]P.-B. Mathayo, D.-K. Kang. Beta and Alpha Regularizers of Mish Activation Functions for Machine Learning Applications in Deep Neural Networks, International Journal of Internet, Broadcasting and Communication 14(1)(2022) 136-141.
[15]K.-M. He, X.-Y. Zhang, S.-Q. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 37(9)(2015) 1904-1916.
[16]D.-H. Wu, S.-C. Lv, M. Jiang, H.-B. Song, Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments, Computers and Electronics in Agriculture, 178(2020) 105742.
[17]A. Womg, M.-J. Shafiee, F. Li, B. Chwyl, Tiny SSD: A tiny single-shot detection deep convolutional neural network for real-time embedded object detection, in: Proc. 2018 IEEE Conference on Computer and Robot Vision (CRV), 2018.
[18]M. Ahmed, R. Seraj, S. Islam, The k-means algorithm: A comprehensive survey and performance evaluation, Electronics 9(8)(2020) 1295.
[19]X.-X. Chu, Z. Tian, Y.-Q. Wang, B. Zhang, H.-B. Ren, X.-L. Wei, H.-X. Xia, C.-H. Shen, Twins: Revisiting the design of spatial attention in vision transformers, Advances in Neural Information Processing Systems 34(2021) 9355-9366.
[20]S.-C. Liu, H. Ma, Combined attention mechanism and CenterNet pedestrian detection algorithm, in: Proc. 2021 IEEE Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2021.
[21]X.-Z. Xie, L. Li, S. Lian, S.-H. Chen, Z.-M. Luo, SERU: A cascaded SE-ResNeXT U-Net for kidney and tumor segmentation, Concurrency and Computation: Practice and Experience 32(14)(2020) 5738.
[22]Z.-L. Zhong, Z.-Q. Lin, R. Bidart, X.-D. Hu, I.-B. Daya, Z.-F. Li, W.-S. Zheng, J. Li, A. Wong, Squeeze-and-Attention Networks for Semantic Segmentation, in: Proc. 2020 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.
[23]Z. Zhu, D. Liang, S.-H. Zhang, X.-L. Huang, B.-L. Li, S.-M. Hu. Traffic-sign detection and classification in the wild, in: Proc. 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[24]Q.-L. Wang, B.-G. Wu, P.-F. Zhu, P.-H. Li, W.-M. Zuo, Q.-H. Hu, ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks, in: Proc. 2020 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020.