Target Detection of Pointer Instrument based on Deep Learning

Yi Wang [a], Guanglin Dong [b]

College of Electrical Engineering, North China University of Science and Technology, Tangshan, China

[a] wangyi@ncst.edu.cn, [b] chengjia@ncst.edu.cn

Abstract. This paper proposed an instrument target detection algorithm based on yolov3 network for the drawbacks caused by manual inspection of pointer instruments in complex industrial environments. Firstly, the algorithm improved the model convergence speed by introducing the k-means++ algorithm to cluster out 9 sets of initial anchor boxes suitable for the pointer meter data set. Moreover, by combining the channel attention mechanism with spatial attention mechanism in the yolov3 backbone network, the extraction of shallow features was further improved by adding two residual blocks to the second residual block, then a new model yolov3-CBAM (Convolutional Block Attention Module) was formed. In addition, the mean average accuracy (map) of the training and testing of the three types of instruments on the data set reaches 90.8% by the results, which is about 2.1% higher than the original yolov3. This algorithm has obvious advantages in the patrol inspection and identification of industrial instruments.

Keywords: Pointer Instrument; yolov3; Attention Mechanism; CBAM; Detection; Feature Extraction.

1. Introduction

Although digital instruments tend to be mature and widely used in data detection, pointer instruments still have an irreplaceable position in the complex environment of strong electromagnetic because of its good stability and durability, such as petrochemical industry, military aerospace, electric power and so on. In the industrial environment, the inspection of instruments still depends on labor. This method not only requires a lot of human resource cost, but also is vulnerable to environmental impact and low work efficiency. Therefore, it is particularly important to identify instruments in a more intelligent way in the field of patrol inspection.

Traditional instrument target detection based on machine vision usually uses artificially designed feature extractors such as Harris corner detection, HOG direction gradient histogram, SVM support vector machine and SIFT feature matching algorithm to extract image features and detect them. For example, Fang Hua et al. [1] used template matching algorithm and sift scale invariant feature transformation algorithm to detect substation instrument targets, which was successfully tested in a 500 kV intelligent substation in China. Zhu Bailin et al. [2] used ORB corner detection algorithm to match pointer instrument images from template library, which basically met the requirements of instrument target detection. However, the detection speed and generalization ability of these traditional instrument target detection algorithms are not ideal, and they are easy to be affected by fog, illumination and other complex environments

Nowadays, artificial intelligence technology is developing rapidly, especially deep learning technology plays a more and more important role in the field of monitoring, target detection, medical image analysis and other practical production and life [3]. It is an important direction in the field of industrial automation instrument detection to apply deep learning technology to instrument automatic reading recognition. Wang Lu et al. [4] adopted the region based Fast RCNN two-stage target detection algorithm for the pointer instrument in the substation environment. The map accuracy on the test set reached 90.2%, but the detection speed was lower than that of the one-stage target detection algorithm. Sun Shunyuan et al. [5] improved the SSD algorithm based on regression, replaced the basic network and added the feature pyramid FPN, which achieved obvious results in the digital instrument test set.

In view of the balance between detection accuracy and speed of Yolo series algorithms, yolov3 is selected as the detection basic network to improve the anchor box calculation method, so as to make it more suitable for the anchor boxes of instrument data set and make it more rapid convergence model; In addition, the residual block is added to further improve the extraction of shallow features, and the convolution block attention module CBAM [6-8] is added to the backbone network to deduce the attention weight sequentially from the two dimensions of channel and space, and then multiply it with the original feature map for adaptive adjustment, so as to make the network pay more attention to the area where the instrument target is located.

2. Yolov3 Target Detection Algorithm

Yolov3 [9] model mainly includes two parts: backbone network darknet53 and Yolo detection layer. The darknet53 backbone network draws lessons from the residual structure of ResNet which is mainly a deep residual network composed of 5 down sampling blocks and 23 residual blocks. Such a deep network can be used as a feature extraction network to extract deeper feature information, so as to obtain a feature map of a specific size.

The feature pyramid is used for prediction in yolov3 detection network. Unlike yolov2, the size transformation of tensor is completed through maximum pooling during forward propagation. Yolov3 adjusts the step size of mask to change the size of feature map. After 5 convolution operations in darknet53, the final output feature map becomes 1/32 of the original, and the input size needs to be set to a multiple of 32, 416 × 416 is selected for this paper and the minimum characteristic diagram of the final output is 13 × 13. In addition, in the middle of the Darknet53 network, 52×52 and 26 × 26 two scale feature maps are extracted after the third and fourth down sampling. After splicing operation, three characteristic images with different scales are finally formed to form Yolo detection layer.

3. Improved Yolov3 Instrument Detection Dlgorithm

3.1 Improved Algorithm for Anchor Box Design

The anchor box mechanism in yolov3 network draws lessons from fast RCNN [10], which is different from fast RCNN. It is a single-stage network, which can directly learn the accurate position of the target by using the anchor box. Based on the predicted feature map, the principle is shown in Figure 2. The main learning parameters of the network are tx, ty, th and tw, which are further adjusted according to the overlapping area ratio with the real frame, and approach the real frame by continuously adjusting the center coordinate offset and width height scaling ratio of the preselected frame; If there is no real dimension box during prediction, but a priori box can give an approximate area in advance and adjust the prediction on this basis. The parameters in the figure meet the following formula:

$$b_x = \sigma(t_x) + C_x \tag{1}$$

$$b_y = \sigma(t_y) + C_y \tag{2}$$

$$b_h = p_h e^{t_h} \tag{3}$$

$$b_w = p_w e^{t_w} \tag{4}$$

Where bx, by, bh and bw are the central coordinates and height and width of the prediction frame; Cx and Cy are the coordinates of the upper left corner of the grid where the center of the prediction frame is located; σ is a sigmoid function that normalizes the coordinates to 0-1 to ensure that the center point is in the grid; pw and ph are preset anchor boxes that map to the width and height in the feature map, so the preset anchor box affects the speed and accuracy of detection.

There is no need to set the anchor box manually. The k-means algorithm can directly calculate the anchor box size required by our own data set. The clustering object is the width and height of the anchor box. Therefore, it is necessary to convert the coordinate information of all annotation boxes

into width and height and extract them in advance, and then the anchor box size can be clustered automatically. The specific process is as follows: firstly, K values are arbitrarily selected in all anchor box data and set as the cluster center; Then, the intersection and union ratio between each real frame and K anchor boxes is calculated, and the real frame is classified to the anchor box with the smallest distance error. Next, the cluster center is updated for the real frame to which each anchor box belongs; Finally, repeat the above steps until the classification of all real boxes does not change.

However, the k-means [11] method randomly selects the value of the initial point K, which has a great impact on the clustering results. Moreover, the initial anchor box of yolov3 is determined according to the coco data set and is not suitable for the instrument data set in this paper. Therefore, this paper adopts the K-means++ clustering [12] algorithm to replace K-means to cluster the anchor box suitable for the instrument data set in this paper. K-means++ algorithm optimizes the selection of initial points, which is not randomly selected. The specific process is as follows: first, randomly select a point as the cluster center; Then calculate the intersection specific distance between each anchor box and the center. The farther the distance, the greater the probability of becoming the cluster center. Repeat the calculation until K centers are found; Finally, nine anchor boxes of different sizes suitable for the instrument data set are calculated according to the K-means clustering process. Finally, the nine anchor boxes of the instrument data set are (115, 92), (151, 114), (173, 149), (194, 195), (231, 175), (267, 213), (315, 237), (363, 281), (463, 342).

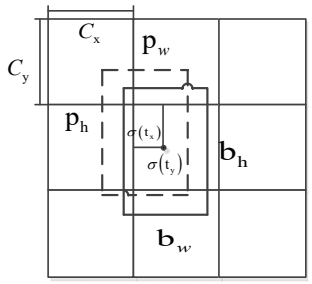


Figure 1. Schematic diagram of prediction frame

3.2 Convolution Block Attention Module CBAM

The attention mechanism [13-15] is similar to human visual attention. In a field of vision, we will first focus on the most important places, and then focus on the secondary parts. At first, the attention mechanism was mainly used for natural language analysis, especially in machine translation scenes. It was widely used to allocate attention by applying different weight sequences to words in the sentence context. In recent years, it has been gradually applied to the image field, which can be easily applied to various CNN networks. Attention mechanism is mainly divided into CAM (channel attention module) and SAM (spatial attention module). Its main functions in the image are to pay attention to what the target elements are and where the target elements are. Unlike the SEnet model, which only applies attention in the channel dimension, this paper uses two attention modules in series to form a convolution block attention module CBAM [16-17], and the structure is shown in Figure 2.

CBAM module exerts attention from the two dimensions of channel and space. CBAM module first compresses the feature map through CAM module, i.e., a global maximum pooled and an average pooled respectively to form two descriptors with a size of 1×1×C. The descriptors are respectively sent to the shared MLP (multi-layer perceptron), and the channel weight coefficient named MC is obtained by summing and compressing two different outputs.

Next, the channel weight coefficient is multiplied by the input characteristic diagram to F1, as shown in equation (5). Then enter the SAM module and pay attention to the target information in spatial location, i.e., two information descriptors with a size of 1×H×W are obtained by maximum pooling and average pooling on the channel. They are combined, and then the spatial weight coefficient named MS is obtained through convolution and scaling. Multiply the MS with the feature map processed by the CAM module to obtain the final feature map F2 with channel and spatial weight, as shown in equation (6)

$$F_1 = M_C(F) \otimes F \tag{5}$$

$$F_2 = M_S(F_1) \otimes F_1 \tag{6}$$

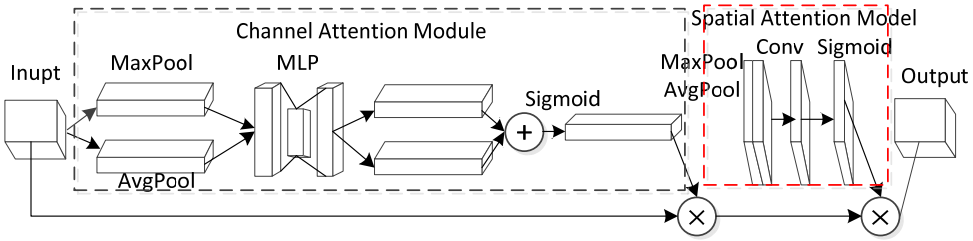


Figure 2. CBAM structure diagram

3.3 Yolov3 Network Structure Improvement

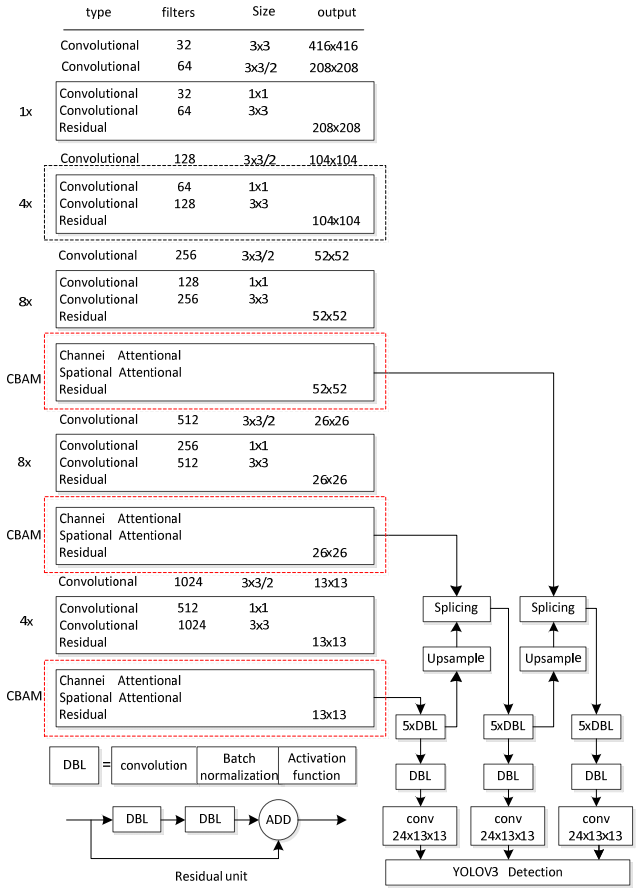


Figure 3. Yolov3-cbam structure

By analyzing the structure and performance of convolution block attention module CBAM, an improved network based on yolov3 is proposed for the instrument data set in this paper, which is called yolov3 CBAM for short. The network structure is shown in Figure 3. On the basis of yolov3 network, there are two main improvements: first, as shown in the red box in Figure 3, CBAM module is added before each output feature map and after the corresponding residual block. By applying attention weight to the feature maps of different sizes extracted by the network from the two dimensions of channel and space, it makes it pay more attention to the target area of pointer instrument in the image, Ignoring other secondary information. The shape of the feature map passing through the CBAM module has not changed, but the importance of different features of the image is divided. Moreover, the attention module CBAM brings fewer parameters, and the amount of calculation can be ignored. The second improvement is that by adding two residual blocks to the second residual

block and changing it into four residual blocks, the network can further extract deeper and more detailed feature information and improve the detection accuracy during instrument detection, as shown in the black dotted box in Figure 3.

4. Experimental Analysis

4.1 Dataset and Network Training

Because there is no public standard pointer instrument data set on the Internet, the data set in this paper is made of field photos taken in pump and valve production plants and laboratories and network crawling photos, including three categories of instruments, a total of 800. The circular meter is labeled as meter, the square meter as square meter and the water meter as water meter. This data set is first manually labeled, and then in order to avoid over-fitting, the data set is enriched to 8,000 by data enhancement operations such as mirror flipping, changing brightness, adding Gaussian noise, and affine transformation. Finally, the xml file with specific annotation information is converted to a txt format file to start training. The experimental running environment is Ubuntu 18.04 5 LTS system, NVIDIA GPU Tesla T4, CUDA 11.2, tensorflow2 0 deep learning framework, python 3.7 interpreter. Firstly, on the constructed data set, the anchor box suitable for the instrument data set is clustered by the improved anchor box clustering algorithm, and then trained in this experimental environment. In addition, the data set is divided into 8:2 ratio of training set and test set. The network super parameter epoch is set to 100 and batch sizes is set to 8. Using Adam optimizer, the initial learning rate is set to 0.001 and the momentum is 0.9.

4.2 Performance Evaluation and Experimental Results

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

In the field of target detection, the evaluation index [18] in the field of information retrieval is used to evaluate the detection performance of a single category, in which the AP value is the area value under the P-R curve with accuracy and recall as the vertical and horizontal coordinate axis, as shown in formula (7). For the calculation of precision (also known as recall), formula (8) is used. For the calculation of recall (also known as precision), use formula (9).

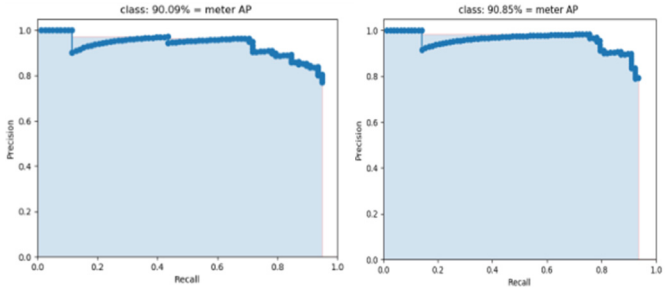
$$AP = \int_0^1 P(r)\, dr \tag{7}$$

$$\text{Preision} = \frac{TP}{TP+FP} \tag{8}$$

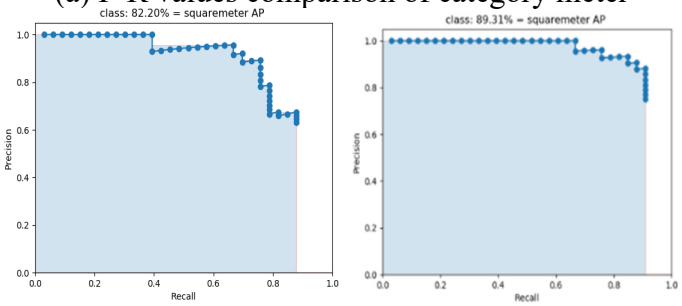
$$\text{Recell} = \frac{TP}{TP+FN} \tag{9}$$

Where TP (real example) means that for a certain type of instrument detection, it is actually an instrument, and the model predicts the number of detection frames of the instrument; FP (false positive example), i.e., the actual area is not the instrument area, but the model considers it to be the number of detection frames in the instrument area; FN (false positive example) is actually the number of instrument areas, but the model does not think it is the number of instrument areas.

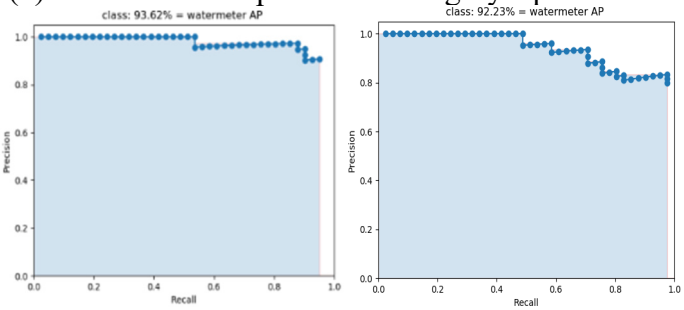
Finally, the weights before and after improvement are obtained after training with the same parameters, and then 200 pictures with a resolution of 600 * 800 are batch tested to obtain the AP value and recall rate of each category of instruments before and after improvement, as shown in Figure 4. Compared with the improved yolov3-CBAM algorithm, the AP value and recall rate of instrument category meter and square meter are improved, and the AP value calculated for category water meter is decreased, but the recall rate is significantly improved.



(a) P-R values comparison of category meter



(b) P-R values comparison of category square meter



(c) P-R values comparison of category water meter

Figure 4. Comparison of P-R values of three types of instruments before and after Yolo improvement

At the same time, for multi category targets, multi category average accuracy (map) is used as the evaluation index, i.e., the AP value of each category is summed and averaged, and the calculation is shown in equation (10). The final map is obtained through the calculation of the above AP values, as shown in Table 1. It can be seen that the improved yolov3-CBAM algorithm improves the map by 2.1 percentage points, reaching 90.8%. The prediction accuracy, small target detection ability and model generalization ability in the actual detection results are better.

$$\text{map} = \frac{\sum_{i=1}^{n} AP_i}{n} \tag{10}$$

Table 1. Comparison Results of Accuracy of Various Types of Instruments Before and After Improvement

index	*Instrument category*	*Before improvement*	*After improvement*
AP/%	meter	90.09	90.85
AP/%	square meter	82.20	89.31
AP/%	water meter	93.62	89.31
map/%	---	88.64	90.80

When the parameters of training data set and training epoch are the same, the weights trained by yolov3 model and yolov3-CBAM model are used to detect the pictures in the test set randomly. The comparison of prediction accuracy is shown in Figure 5. It can be seen that the detection confidence

of yolov3 without improvement is 0.75 and 0.94 respectively, while the detection confidence of yolov3-CBAM algorithm proposed is 0.97 and 0.99. The detection accuracy of the improved algorithm is higher than that before improvement. At the same time, compared with the model before the improvement, the improved model also improves the ability to detect small targets. As shown in Figure 6, the improved algorithm can also detect slightly smaller and occluded target instruments, while the yolov3 detection effect without improvement is not detected. In addition, yolov3-CBAM target detection algorithm has stronger generalization ability. After training, photos are taken from other places, and this picture has never participated in training in any form of data. The comparison of detection effects before and after improvement is shown in Figure 7. It can be seen that yolov3 without improvement detects only one instrument, and the confidence and detection frame are inaccurate, Yolov3-CBAM detected the two instruments, and the confidence reached 0.97 and 0.98 respectively. Finally, the improved algorithm can show good results in detecting instruments with fog and uneven illumination. The detection effect is shown in Figure 8.

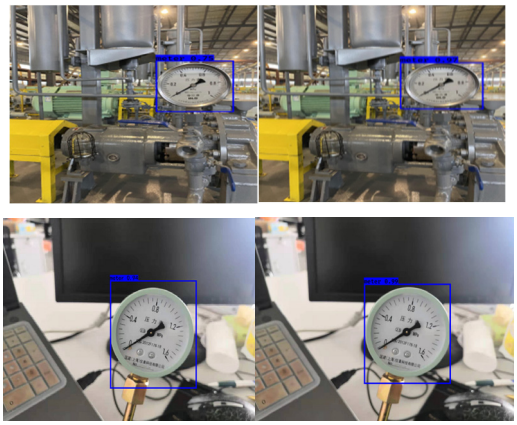


Figure 5. Comparison of detection accuracy effect before and after improvement

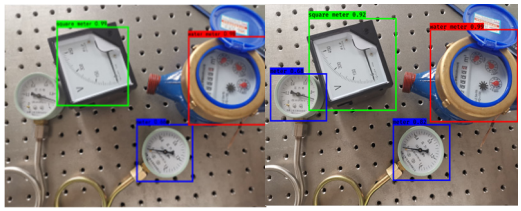


Figure 6. Comparison of small target detection effect before and after improvement

Figure 7. Comparison of generalization ability test results before and after improvement

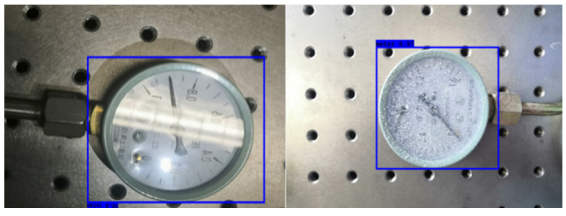


Figure 8. Detection effect of uneven illumination and fog instrument

At the same time, the detection speed is also counted. 200 pictures are detected. The time-consuming of model detection before and after improvement is 28.7s and 31.5s respectively. In addition, under the same data set and test set, it is also compared with fast RCNN and SSD network models. The specific index results are shown in Table 2. It can be seen that the detection speed of the improved yolov3-CBAM algorithm is slightly lower than that of the improved yolov3-CBAM algorithm, but the map value of the model is the highest, reaching 90.8%, which is excellent in the actual experimental detection effect.

Table 2. Performance Comparison of Detection Results of Different Networks

network model	*map/%*	*Time/s*
Faster RCNN	88.96	0.723
SSD	87.51	0.131
Yolov3	88.64	0.144
Paper method	90.80	0.158

5. Conclusion

This paper proposes an improved instrument target detection algorithm yolov3-CBAM based on yolov3. By improving the anchor box calculation method, the attention mechanism module CBAM is embedded in yolov3 network and the residual module is added to further improve the efficiency of feature extraction and detection accuracy. The experimental results show that the yolov3-CBAM instrument target detection algorithm map reaches 90.80%, which is 2.1% higher than the original yolov3 detection algorithm, the recall rate and confidence are also improved, the detection speed is slightly reduced, but the difference is very small, and the algorithm has strong generalization ability. Without training for new instruments or a little training based on this weight can achieve better results, The instrument with uneven illumination and fog can also be detected correctly, which has a good application prospect in the field of instrument automatic inspection. In addition, this paper considers the identification of instruments without reading processing, and the subsequent work needs to read the identified instruments.

Acknowledgments

Tangshan Science and Technology Planning Project (21130212C).

References

[1] Fang Hua, Jiang Tao, Li Hongyu, et al. A recognition algorithm of double needle instrument reading suitable for intelligent substation inspection robot [J] Shandong Electric Power,2013(03):9-13+69.

[2] Zhu Bailin, Guo Liang, Wu Qingwen. Intelligent reading method of pointer instrument based on orb and improved Hough transform [J] Instrument technology and sensors, 2017(01):29-33+73.

[3] Li Xinye, Zhu Jing, Ma Lina. Overview of scene recognition methods based on deep learning [J] Computer engineering and application, 2020, 56 (05): 25-3.

[4] WANG L, WANG P, WU LH, et al.Computer Vision Based Automatic Recognition of Pointer Instruments: Data Set Optimization and Reading.[J]. Entropy (Basel, Switzerland) , 2021, 23(3):272-293.

[5] Sun Shunyuan, Yang ting. Instrument target detection algorithm based on deep learning [J] Instrument technology and sensors, 2021 (06): 104-108.

[6] Bo Jingwen, Zhang Chuntang, fan Chunling, etc Improved detection method of sundries on ore conveyor belt of yolov3 [J / OL] Computer engineering and application: 1-10[2021-07-29] http:// kns. cnki. net/ kcms/ detail/11.2127.TP.2021 0705.0854002.html.

[7] CALUSEN H, GROV G, ASPINALL D. CBAM: A Contextual Model for Network Anomaly Detection [J]. Computers, 2021, 10(6):79-96.

[8] WOO S, PARK J, LEE J Y, et al.CBAM: convolutional block attention module[C]// The European Conference on Computer Vision, 2018: 3-19.

[9] REDMON J, FARHADI A YOLOv3:An Incrementa1 Improvement[C]// IEEE Conference on Computer Visionand Pattern Recognition. Piscataw ay: IEEE Press, 2018:1-6.

[10] GIRSHICK R. Fast R-CNN [C]/ /2015 IEEE International Conference on Computer Vision (ICCV). NewYork, USA:IEEE, 2015. 1440-1448.

[11] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2016:770-778.

[12] Cong Mou, Zhang Ping, Wang Ning Armored vehicle detection method based on improved yolov3 Journal of ordnance and equipment engineering [J], 2021,42 (04): 258-262.

[13] Ruan Chen, Guo Hao, anjubai SAR nearshore ship detection in complex background [J] Chinese Journal of image and graphics, 2021,26 (05): 1058-1066.

[14] Niu Zhaoyang,Zhong Guoqiang,Yu Hui. A review on the attention mechanism of deep learning[J]. Neurocomputing,2021:48-62.

[15] Yang, Zhuoqun, Zhang Tao, Yang, Jie. Research on classification algorithms for attention mechanism [C]//Proceedings of 2020 19th International Symposium on Distributed Computing and Applications for Business Engineering and Science(DCABES). 2020:194-197.

[16] Wang Meihua, Wu Zhenxin, Zhou Zuguang Research on fine-grained identification of crop diseases and pests based on attention improved CBAM [J] Journal of agricultural machinery, 2021, 52 (04): 239-247.

[17] Wan Jialong, Jin Weidong, Tang Peng, etc Retinal vessel segmentation based on visual attention enhancement cbam-u-net model [J] Computer application research, 2020, 37 (S2): 321-323.

[18] ZHOU Y, LIU L, SHAO L, et al. Fast automatic vehicleannotation for urban traffic surveillance [J]. IEEE Transac-tions on Intelligent Transportation Systems, 2018, 19(6):1973-1984.