International Core Journal of Engineering Volume 8 Issue 12, 2022
ISSN: 2414-1895 DOI: 10.6919/ICJE.202212_8(12).0035

Improved Surface Defect Detection of YOLOV5 Aluminum Profiles based on CBAM and BiFPN

Fan Hua

School of Tianjin University of Technology and Education, Tianjin 300000, China

*499227148@qq.com

Abstract

Aluminum profile is one of the most commonly used profiles in daily life. It is widely used in aviation, construction, industry, home decoration, doors and Windows, etc. Because of its huge demand, the production and sales of aluminum profile in China are increasing year by year. Due to the influence of various factors such as mechanical friction, product process and raw materials, the aluminum profiles produced will have different types and degrees of defects. In this paper, a fishing boat target detection model based on improved YoloV5 is proposed. In view of the problems of Convolutional Block Attention Module, convolutional block attention Module is added to the convolutional block attention module. CBAM and weighted Bidirectional Feature Pyramid Network (BiFPN). The results show that the BiFPN feature fusion network combined with cbam improves the mAP of the original YOLOv5 by 5.6%, which effectively enhances the overall performance of network detection.

Keywords

Surface Defect Detection; YOLOV5; CBAM Attention Mechanism; BiFPN Feature Fusion Network.

1. Introduction

In the manufacturing process of aluminum profile, due to the processing method, abnormal conditions of equipment, production of raw materials and many other reasons, the surface of aluminum profile will appear bubbles, convex powder, orange peel, scratches and other defects, seriously affect the quality of aluminum profile products. Therefore, it is of great significance to detect the surface quality of aluminum profiles and repair the defective products in time.

In the process of industrial production, there are three main methods used to detect surface defects [1] : artificial naked eye detection, artificial physical detection and machine vision-based detection. The first is to identify surface defects by artificial eyes, which is commonly used by some small and medium-sized enterprises. This method mainly relies on human eyes for recognition. When human eyes are tired, identification errors may occur, so there are problems such as low detection efficiency, subjective error and contact damage. The second method is to detect by some physical methods, typical detection methods mainly include: Resonance Ultrasonic Vibrations (RUV) method, electronic speckle interference analysis, etc. This method based on artificial physics detects problems such as expensive equipment, complex operation and limited scope of use, so many manufacturers do not choose this method. With the continuous development of manufacturing automation and people's quality of life, people's demand for product quality has become more and more great. This puts forward higher requirements for defect detection technology, while the traditional detection by artificial naked eye not only has a long working time, but also may be faced with harsh environment, which is the reason for the low detection accuracy and efficiency of this method. The last is through

ISSN: 2414-1895 DOI: 10.6919/ICJE.202212_8(12).0035

machine vision to detect, this method is different from the first method, it relies on computer image processing for target detection, mainly has the advantages of low cost, detection quasi, detection of non-destructive, simple equipment and so on. Through the defect detection method based on machine vision to detect the products produced in industry has two main advantages, which are to reduce the cost and improve the speed of production. With these two advantages, the detection method based on machine vision has still become the mainstream method of industrial defect detection.

The technology of object surface defect detection originated in the 20th century and has gone through four stages, namely manual detection, physical property detection, machine vision-based defect detection and deep learning-based defect detection. Compared with China, foreign research on object surface defect detection technology started earlier. In the late 20th century, it has been gradually introduced into the actual production of enterprises.

In foreign countries, the technology of industrial defect detection has been applied to industrial production in the 1970s. Literature 2 studies the cold-rolled strip steel and proposes an online detection method for the quantity, category and quality level of surface defects of cold-rolled strip steel through artificial neural network. In Reference 3, an advanced lighting system is adopted in image acquisition, and then the surface defects of metals are detected by an image processing algorithm. The experimental results show that this algorithm can improve the recognition rate. Reference 4 proposed a combination of lighting pipeline and charge-coupled element linear array camera to detect surface defects of steel plate. Reference 5 proposes an image processing technology that combines multiple filtering algorithms and templates to automatically detect objects.

At present, China is also in continuous development and has made great achievements in surface defect detection. In reference 6, a method based on machine vision is proposed to detect the defects on the cable surface. In the image acquisition, advanced lighting source is used to ensure the quality of the image. The experimental results show that the defects of various cables can be detected, and the false detection rate has been greatly improved. Reference 7 takes rail surface as the research object and proposes an algorithm combining dynamic threshold segmentation and region extraction. Firstly, the defect image of rail surface is processed, and then the location area of the defect is found for extraction. Reference 8 puts forward a detection algorithm based on image processing for the surface defects of solar cells, and realizes the detection of solar cells through contour analysis, dynamic threshold segmentation and other methods. Experiments show that it has good performance.

In this paper, YOLOv5s is selected as the pre-training model, and the base transfer learning is applied to the detection of steel surface defects for the purpose of protection.

The accuracy of the model is further improved under the premise of verifying the detection speed. Firstly, the attention mechanism is added to improve the channel and space connection of target features, enhance the network's attention to the key information in the feature map, and facilitate the network to extract and utilize features more perfectly. The Feature fusion Network of the original YOLOv5 model is Path Aggregation Network (PANet), although compared with the Feature Pyramid Network (FPN), it can fuse the features of targets at different scales better, thus improving the effect. But there is room for improvement. Therefore, this paper goes further in the direction of FPN and tries a more complex bidirectional fusion, BiFPN. BiFPN is a weighted bidirectional feature pyramid network that allows simple and fast multi-scale feature fusion. The author's aim is to pursue a more efficient multi-scale fusion method. In the past feature fusion, the BiFPN introduced weight, which can better balance the feature information of different scales.

2. The YOLOV5 Algorithm

2.1 Network Structure

There are four versions of the YOLOV5 model, which are YOLOV5s, YOLOV5l, YOLOV5 and YOLOV5x. The network structure of the four models is the same, and the difference lies in the depth coefficient depth_multiple and width coefficient width_multiple.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202212_8(12).0035

Taking YOLOV5s as an example, Figure 1 shows its network structure.

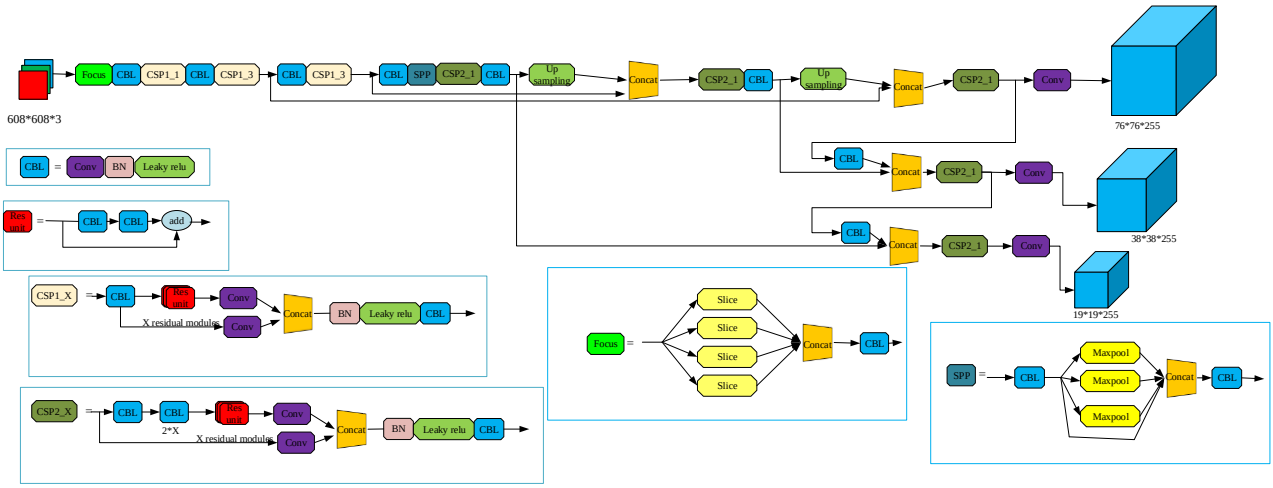


Figure 1. YOLOV5s network structure

YOLOV5s target detection process is as follows:

First of all, the input end will scale the image to get a picture of 608*608 pixels.

The backbone Network uses Cross Stage Partial Network (CSP) [9] for image feature extraction, and the neck network uses FPN+PAN for information combination like YOLOV4.

Detection head is responsible for detecting the category and position information of objects.

YOLOV5 has the following five main components:

Focus: includes 4 slicing modules;

CBL: It is composed of convolution layer, BN layer and Relu activation function;

Res Unit: Using the residual structure of Resnet network for reference, a deeper network can be built;

CSP1-X: consists of 1 CBL and X Res unint modules and 2 convolution layers;

CSP2-X: consists of 2X CBL sum and 2 convolution layers;

SPP: consists of 1 CBL and 3 maximum pooling layers.

Concat: tensor concatenation, which expands the dimension of two tensors.

2.2 Evaluation Indicators

1) Accuracy rate and recall rate

The accuracy rate is the percentage of all predictions that are positive that are positive. Recall rate refers to the proportion of all predicted positive categories that are actually positive in the total positive category. The formula is as follows:

$$R_p = \frac{T_P}{F_P + T_P} \tag{1}$$

$$R_R = \frac{T_P}{F_N + T_P} \tag{2}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202212_8(12).0035

Where, R_p is the accuracy rate, R_R is the recall rate. T_P (True Positive) is the number of samples labeled as Positive and predicted as positive; F_P (False Positive) is the number of samples labeled as Negative but predicted as positive incorrectly; FN (False Negative) is the number of samples labeled as positive but predicted as negative incorrectly.

2) Average accuracy and average accuracy mean

Average Precision (AP) is the most basic evaluation index in target detection tasks. Its formula is as follows:

$$AP = \int_0^1 R_p(t)dt \tag{3}$$

Where, $R_p(t)$ represents the accurate rate when the threshold is taken. When a target detection tasks include multiple categories, usually using the Average Precision Average (Mean business, Precision, mAP) to reflect the accuracy of all categories. The formula is as follows:

$$mAP = \frac{\sum_{n=1}^{N} AP_n}{N} \tag{4}$$

Where, N is the number of object categories, n represents the first category, and AP_n represents the detection accuracy of the network model for objects of the NTH category.

Another important metric in the context of industrial testing is detection speed, which is the number of frames or images detected within 1s.

3) Loss function

The loss function of Yolov5 is divided into three parts: classification loss function, confidence loss function and frame positioning loss function. The final loss function is the sum of the three parts. The classification loss function and confidence loss function are binary cross entropy loss. The frame positioning loss function uses CIOU loss. IOU is the intersection ratio, used to evaluate the distance between the predicted box and the real box. IOU alone cannot measure the effect of target box regression. When two border steps overlap, the IOU is 0, which does not reflect the distance between the two borders, and a gradient of 0 cannot be optimized. The Distance Intersection over Union (DIoU) improved it, taking into account the changes of distance, overlap rate and scale between the target frame and anchor. DIOU can be calculated from Equation 2.5, where b and bgt are the center points of the prediction and real boxes respectively. C is the diagonal of the minimum closure region of the prediction box and the real box. b and b^gt represent the center points of the prediction and target boxes, respectively. ρ^2 calculates the Euclidean distance between two points. DIOU can calculate the gradient even if the two borders overlap step by step. Equation 2.6 is the loss function based on CIOU. CIOU further improves on DIOU, taking into account the aspect ratio of the bezel. V represents the ratio similarity between the length and width of the frame, and α is the weight function.

$$DIOU = IOU - \frac{(\rho^2(b, b^g t))}{c^2} \tag{5}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202212_8(12).0035

$$\begin{cases} L_{CIOU} = 1 - IOU + \dfrac{\rho^2(b, b^{gt})}{c^2} + \alpha v \\ v = \dfrac{4}{\pi^2}(arctan\dfrac{w^{gt}}{h^{gt}} - arctan\dfrac{w}{h})^2 \\ \alpha = \dfrac{v}{(1 - IOU) + v} \end{cases} \tag{6}$$

3. Network Structure Improvement

3.1 CBAM

CBAM[10-13] is a lightweight attention module that can simultaneously conduct attention mechanism in space and channel. FIG. 2 shows the structure of CBAM's attention mechanism. It mainly consists of two modules, namely channel attention module and spatial attention module, which adopt the form of series.

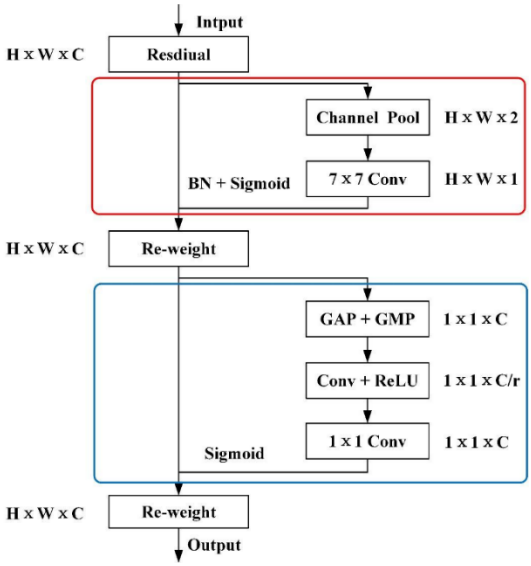


Figure 2. CBAM structure diagram

3.1.1 Channel Attention Module

Channel attention It makes sense to focus on which channel to focus on, as follows:

Firstly, global average pooling and global maximum maximum pooling are carried out for input features;

Then output the weight of the channel's attention through the fully connected neural network of two layers;

The standard weight coefficient of each channel is obtained by sigmiod function;

Finally, each weight is added to the weight of the original channel to reclassify the importance of different information on the number of channels in the initial feature map.

The process of the channel attention mechanism can be represented by formula 7.

$$\begin{aligned} M_c(F) &= \sigma\left(MLP(AvgPool(F) + MLP(MaxPool(F)\right) \\ &= \sigma(W_1(W_0(F_{avg}^c) + W_1(W_0(F_{\max}^c)) \end{aligned} \tag{7}$$

Where, $M_c(F)$ is a one-dimensional attention feature graph, σ is the activation function of sigmiod, MLP is a multi-layer perceptron composed of two fully connected neural networks,

ISSN: 2414-1895 DOI: 10.6919/ICJE.202212_8(12).0035

$AvgPool(F)$ and $MaxPool(F)$ is the result of the feature graph through GAP and GMP, W_1 and W_0 is the two weights of MLP, F_{avg}^{c} and F_{avg}^{c} is the corresponding one-dimensional mapping of $AvgPool(F)$ and $MaxPool(F)$.

3.1.2 Spatial Attention Module

After the output of channel attention, the spatial attention mechanism is introduced to focus on the spatial feature information. The specific steps are as follows:

The pixel size of the input image or feature map compressed by the pooling layer according to the pixel size;

Compress the convolution kernel and Relu activation function of the compressed feature graph again;

The convolution check is used for upsampling of the compressed feature map, which conforms to the pixel size of the input of the next layer;

Combine the space and channel attention modules to get the feature map, and finally conduct the secondary division of the information on the space and channel.

The process of the spatial attention module can be represented by formula 8.

$$\begin{aligned} M_c(F) &= \sigma(f^{7\times7}([AvgPool(F); MaxPool(F)])) \\ &= \sigma(f^{7\times7}([F_{avg}^{s}; F_{\max}^{s}])) \end{aligned} \tag{8}$$

Where, $f^{7\times7}$ is the convolution operation of filter size is 7×7, F_{avg}^{s} and $F_{\max}^{s}$ is the two-dimensional mapping of average pooling feature and maximum pooling feature respectively.

CBAM performs pooling operations in the feature graph in order to compress the spatial features. The convolution kernel of 7×7, greatly reduces the parameters and computation. The two convolution operations not only maintain the correlation of spatial features, but also ensure that the input and output do not change. After the feature map is processed by CBAM, the dual attention weight of channel and space can better extract key features.

3.2 BiFPN

Bidirectional Feature Pyramid Network (BiFPN)[14-15] is produced by google. It has two main ideas, namely efficient cross-scale bidirectional connection and weighted feature fusion.

3.2.1 Cross-scale Connection

Feature fusion, in essence, is to input a set of features $P^{\text{out}} = (P_{l_1}^{in}, P_{l_2}^{in}, \ldots)$ and find a transformation function to output the new features $P^{\text{out}} = (P_{l_1}^{in}, P_{l_2}^{in}, \ldots)$.

Figure 3 shows how FPN works, using a top-down path.

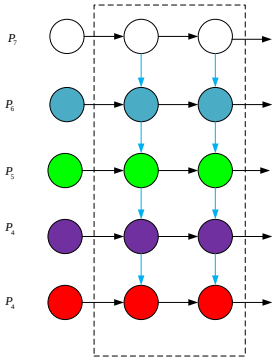


Figure 3. Structure of FPN

ISSN: 2414-1895 DOI: 10.6919/ICJE.202212_8(12).0035

Compared with the top-down path of FPN, PANet adds a bottom-up path on this basis. The structure diagram is shown in 4. By comparison, PANet achieves better results than FPN.

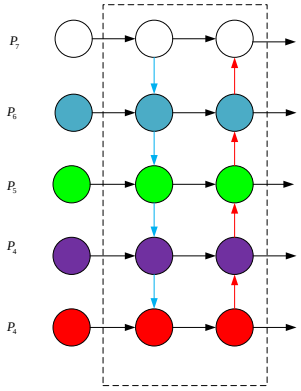


Figure 4. Structure of PANet

BiFPN has the following changes on the basis of PANet:

Delete nodes with only one input. If a node has only one input and no feature fusion, its contribution to feature fusion is small. The change process is shown in Figure 5.

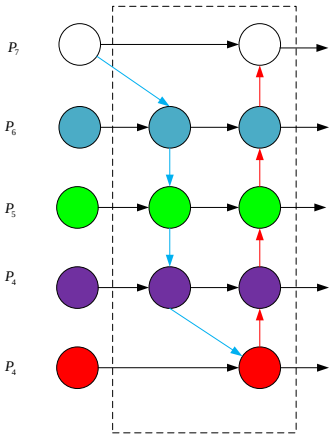


Figure 5. Node deletion diagram

Add jump connections. If the original input is at the same level as the input, add a connection line between the two nodes to get more feature fusion at no additional cost. This is the basic BiFPN unit, as shown in Figure 6.

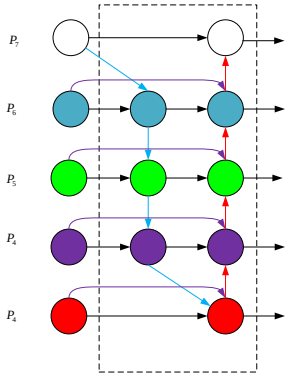


Figure 6. Basic BiFPN unit

ISSN: 2414-1895 DOI: 10.6919/ICJE.202212_8(12).0035

Double stack BiFPN basic unit to get more feature fusion.

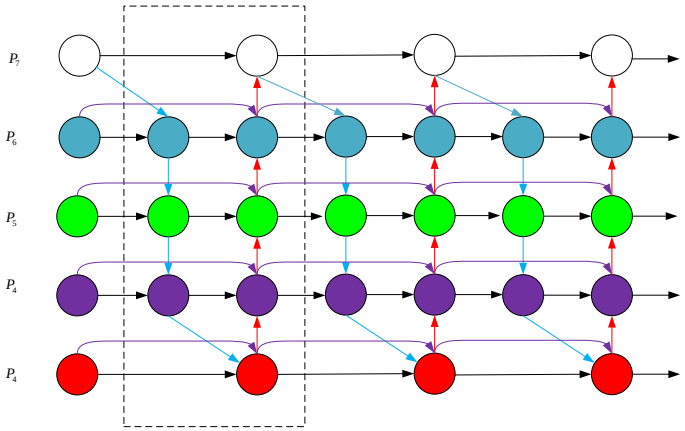


Figure 7. BiFPN network structure

3.2.2 Weighted Fusion

Different features have different resolution and different contribution to feature fusion. BiFPN proposed to add weight, can be a good balance of different scales of feature information. There are mainly three weighting methods:

A) Fusion without boundaries

The only disadvantage of relying on scalars to achieve good accuracy at very low computational cost is that the training is unstable, since scalars are unbounded.

$$0 = \sum_i W_I I_I \tag{9}$$

B) Fusion method based on Softmax

The scope to which ownership is reunified represents the importance of each input.

$$0 = \sum_i \frac{e^{W_i}}{\sum_j e^{W_j}} \tag{10}$$

C)Rapid normalization and fusion

Make sure each weight is non-negative according to Relu, and add a very small one to ensure the stability of the value. The advantage is that it is faster.

$$0 = \sum_i \frac{e^{W_i}}{\varepsilon + \sum_j e^{W_j}} \tag{11}$$

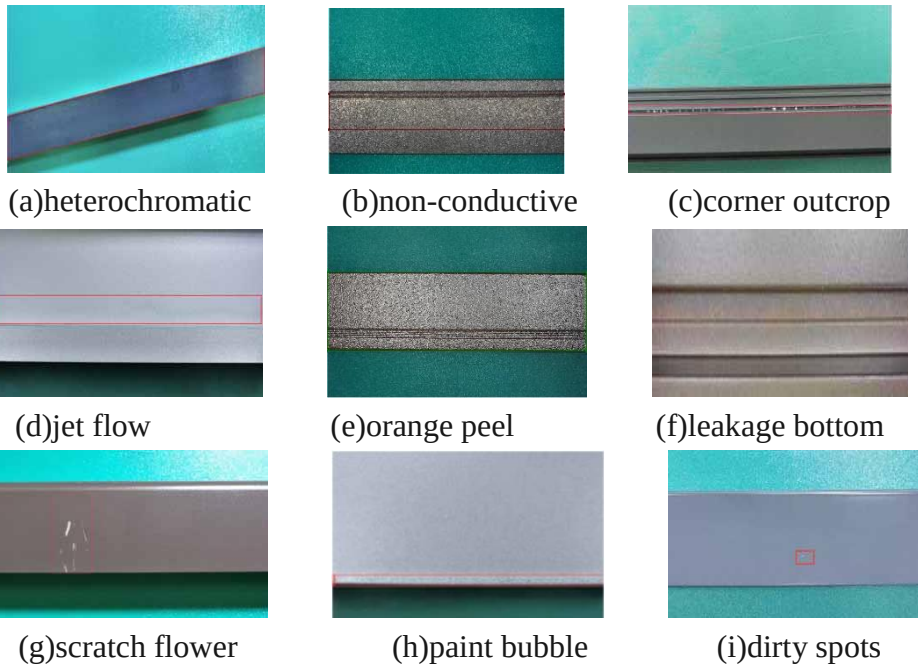
4. Experiment and Result Analysis

4.1 Data Set

The aluminum profile data set is the data collected by Tianchi platform from an aluminum profile manufacturer in Guangdong Province in monitoring the surface defects of aluminum profiles in a

ISSN: 2414-1895 DOI: 10.6919/ICJE.202212_8(12).0035

certain period of time. There are nine types of defect data, which are respectively heterochromatic, non-conductive, corner outcrop, jet flow, orange peel, leakage bottom, scratch flower, paint bubble and dirty spots. The size of the pictures is uniform. Figure 8 shows the sample size of the various defects.



(a)heterochromatic (b)non-conductive (c)corner outcrop

(d)jet flow (e)orange peel (f)leakage bottom

(g)scratch flower (h)paint bubble (i)dirty spots

Figure 8. Surface defect image of aluminum profile

4.2 Experimental Environment Construction

The hardware configuration of the training model in this chapter is the artificial intelligence all-in-one machine of Tianjin Huada Technology. The specific hardware configuration is shown in Table 1. The operating system is Ubuntu16, the deep learning framework is pytorch, the programming language is python3.9, and the GPU acceleration library is cuda7.0.5.

Table 1. Hardware configuration

Name	Model Number
processor	CPU-i5
memory	16g
Solid state memory	250g
The graphics card	NVIDIA RTX2070S
Acceleration module	Cuda9.0/cudacnn7.0.5

4.3 Comparison Experiment and Results

This study mainly improved the YOLOV5 algorithm from two parts, and each part proposed two methods respectively. The first part is to add CBAM attention mechanism in the stage of backbone network feature extraction, and the second part is to replace the feature fusion network of YOLOV5 with BiFPN.

The training results on the aluminum profile data set are shown in Table 2. It can be seen from the table that the model with CBAM attention mechanism and BiFPN feature fusion network has reached the maximum value of 0.722.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202212_8(12).0035

Table 2. Training results

Model	R_P	R_R	mAP
YOLOV5	0.745	0.674	0.666
YOLOV5+CBAM+BiFPN	0.761	0.73	0.722

4.4 Ablation Experiment

In order to verify the effectiveness of the proposed method, ablation experiments were performed on the test set, as shown in Table 3:

Table 3. Comparison of ablation experiments

Method	R_P	R_R	mAP
YOLOV5s	0.745	0.674	0.666
+CBAM	0.735	0.692	0.691
+BiFPN	0.761	0.73	0.722

The first row in Table 3 is the initial YOLOV5 model, the second row adds the CBAM attention mechanism, the accuracy is lower than the initial model, and both the accuracy and mAP are greatly improved, in which the mAP value is increased by 0.3 points. The third row adds the BiFPN feature fusion network after the CBAM attention mechanism is added. BiFPN introduces weight, which can better balance the feature information of different scales. The accuracy rate, recall rate and mAP of the improved algorithm have been improved by 0.024,0.056 and 0.056 points respectively.

5. Conclusion

Based on YOLOV5, the surface defects of aluminum profiles are studied in this paper. The feature fusion network of the original YOLOV5 is improved, and the attention mechanism and feature fusion network are added to improve the detection algorithm. There are two main improvements in the network structure: 1) The addition of attention mechanism can improve the connection between the channel and space of the target features, so as to enhance the network's attention to the key information in the feature map, which is conducive to the network's more perfect extraction and utilization of features. 2) Modify the feature fusion network to BiFPN, which allows simple and fast multi-scale feature fusion. In the past feature fusion, the BiFPN introduced weight, which can better balance the feature information of different scales.

References

[1] LIU Lizhe. Research on Surface Defect Detection Algorithm Based on Deep Learning [D]. Huazhong University of Science and Technology,2019.

[2] Li Zhifeng. Surface Defect Detection System of galvanized strip Steel Based on Image Recognition [D]. Dalian: Dalian University of Technology, 2018.

[3] Guo H R. Research of defect recognition method for highlight rotating surface[D]. Xi'an: Xi'an University of Technology, 2018.

[4] Wang L. 3D detection of surface of metal strips and plates based on photometric stereo[D]. Beijing: University of Science and Technology Beijing, 2019.

[5] Yang Shuman. Research on Crack Detection System of Precision Castings Based on Image Processing [D]. Taiyuan: Taiyuan University of Science and Technology, 2017.

[6] Zhang Jun. Research on Online Detection System of Cable Apparent Defects Based on Machine Vision [D]. Chengdu: University of Electronic Science and Technology of China, 2017.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202212_8(12).0035

[7] Liu Ze, Wang Wei, Wang Ping. Design of Machine Vision System for Rail Surface Defect Detection [J]. Journal of Electronic Measurement and Instrument, 2010,24 (11) : 1012-1017.

[8] Zhou Qi. Design of solar cell Defect Detection System based on HALCON [D]. Zhenjiang: Jiangsu University, 2017.

[9] Lin Tsung-Yi, Goyal Priya, Girshick Ross, et al. Focal Loss for Dense Object Detection.[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.

[10] Woo S, Park J, Lee JY, et al. CBAM: convolutional block attention module[C]//Proceedings of The European Conference on Computer Vision (ECCV). 2018: 3-19.

[11] Chen Huasuo, Jiao Ge. Deep Learning Image Steganalysis Method Fused with CBAM[C]. 2022, Proceedings of the 12th International Conference on Computer Engineering and Networks:1175-1184.

[12] Li Zhiyong, Jiang Xueqin, Shuai Luyu, et al. A Real-Time Detection Algorithm for Sweet Cherry Fruit Maturity Based on YOLOX in the Natural Environment[J]. Agronomy, 2022, 12(10):2482-2482.

[13] Xue Zhenyang, Lin Haifeng, Wang Fang, et al. A Small Target Forest Fire Detection Model Based on YOLOv5 Improvement[J]. Forests, 2022, 13(8):1332-1332.

[14] Tan M, Pang R, Le Q-V.Efficient Det: Scalable and Efficient ObjectDetection[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Seattle, WA, USA: IEEE, 2020: 0778-10787.

[15] Guo Shuyi, Li Lulu, Guo Tianyou, et al. Research on Mask-Wearing Detection Algorithm Based on Improved YOLOv5[J]. Sensors, 2022, 22(13):4933-4933.