ISSN 2616-5775 Vol. 6, Issue 3: 92-101, DOI: 10.25236/AJCIS.2023.060312

A biscit defect detection method based on improved YOLOv5

Shulin Li[1], Hong Zhao[1,*], Yuxi Huang[1]

[1]*College of Mechanic & Control Engineer, Guilin University of Technology, Guilin, China*
[*]*Corresponding author*

Abstract: *Defect detection is of great importance to ensure the quality of biscuit production. An improved YOLOv5 biscuit detection algorithm is proposed for the problems of poor real-time and low accuracy of biscuit defect detection methods. First, the number of C3s in the backbone network is reduced, and then the depth-separable convolution is used instead of the normal convolution in the network to reduce the model parameters and computation and improve the detection speed. Secondly, the SE attention module is added to the feature extraction layer to enhance the feature extraction capability of the backbone network and improve the accuracy of biscuit defect detection. Finally, the EIOU loss function is introduced to accelerate the model convergence and accurate target localization. The improved algorithm is tested on the self-built biscuit dataset, and the experimental results show that: the detection accuracy of the proposed algorithm can reach 99.2%, and the detection speed is 67 frames/s, which can meet the actual production requirements.*

Keywords: *biscuit defect, YOLOv5, lightweight, attention mechanism*

1. Introduction

In industrial production, surface defects on objects are a visual reflection of product quality problems. To ensure the quality of products in automated production, object surface defect detection has become increasingly important. In recent years, deep learning-based defect detection has become the mainstream method in the field of defect detection, which treats defects as objects and uses convolutional networks to extract abstract and powerfully representative features, and then locates and classifies defects [1].

Depending on whether the convolutional network contains candidate frames, defect detection algorithms can be further divided into two types: a two-stage model based on region suggestion represented by Faster R-CNN [2]. Although this type of algorithm has high detection accuracy, the number of model parameters is large, and the network training and detection speed is slow, which is not suitable for practical projects. Another type of algorithm is the one-stage model without region suggestion represented by SSD [3] and YOLO [4], which has received a lot of attention because it greatly improves the detection speed with a slight loss of detection accuracy. However, YOLO algorithms still have some problems in practical applications, such as the contradiction in detection speed and accuracy, and the poor detection of small targets. Therefore, researchers have improved the YOLO algorithm to enhance its performance in defect detection tasks. For example, Wang [5] proposed a method to reduce the number of convolutional parameters of the YOLOv5 model using depth-separable convolution to improve the computational efficiency. Ma [6] combined depth-separable convolution and a two-channel attention mechanism to design a lightweight convolutional block that can greatly reduce the computational complexity and scale of the model while ensuring the detection accuracy.

He [7] proposed an innovative joint attention layer that combines the advantages of the SE module and the spatial attention SAM module to extract useful feature information more accurately and improve the accuracy of defect detection. Zheng [8] proposed an improved YOLOv5 model by using the SE module in the backbone network of YOLOv5 and using the ACON activation function replaces the Leaky ReLU activation function in the CSP part to improve the robustness and generalization of the model and to improve the detection accuracy. Due to the large number of deep learning network parameters and high computational resource consumption, the biscuit defect detection system requires lightweight improvements to improve computational performance. In addition, the relative density of biscuits on the conveyor belt and their mutual occlusion can lead to difficulty in feature extraction, and the lightweight model can easily lead to degradation of detection accuracy. To this end, an improved biscuit defect detection algorithm is proposed in this paper, which uses reducing the number of convolutional layers in

ISSN 2616-5775 Vol. 6, Issue 3: 92-101, DOI: 10.25236/AJCIS.2023.060312

the network, introducing a lightweight convolutional module, and incorporating an attention mechanism to improve the deployment efficiency and accuracy of the model. Meanwhile, a new loss function EIOU is used to reduce the error.

2. YOLOv5s Model

YOLOv5 is one of the best versions of the performance table in the YOLO series, and as the depth and width of the network increase, it is divided into four different models: YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x [9]. Among them, YOLOv5s has the least number of network parameters, is easier to deploy, and has high detection accuracy to meet the needs of industrial production. Therefore, YOLOv5s is selected as the basic model of this paper.

The network structure of the YOLOv5 model consists of four parts, namely the input layer, the backbone, the neck layer and the output layer. The network structure is shown in Figure 1.

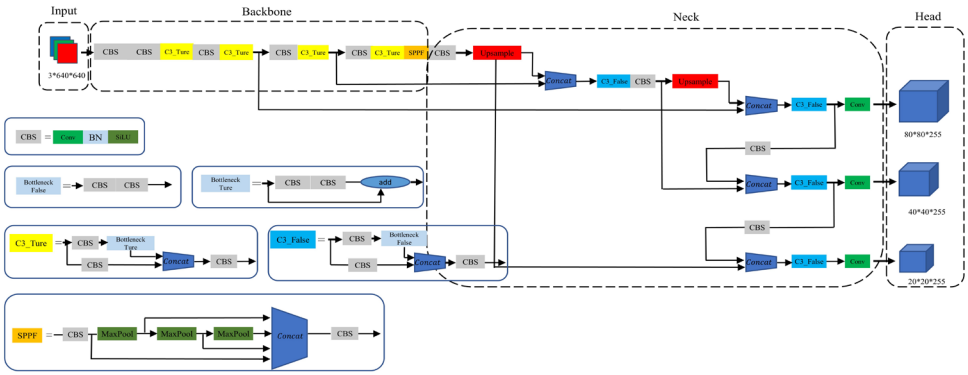


Figure 1: YOLOv5s network structure

The input layer mainly includes mosaic data augmentation, image size processing and adaptive anchor box calculation. Mosaic data augmentation enriches the image background and the number of small targets by randomly cropping and stitching four images.

The backbone layer includes CBS modules, C3 modules and SPPF modules. The CBS module is the basic component in the YOLOv5 network structure, which is composed of Conv+Bn+SiLU activation function. The C3 module is composed of a CBS module and a residual module, and C3_True applied to the backbone layer, which is a deep network whose residual structure can increase the gradient value of backpropagation and avoid the disappearance of the gradient. The SPPF module is an improvement on SPP, serializing the input through the maximum pooling layer of the same core size and then integrating features.

The neck layer adopts FPN+PAN structure. From top to down, FPN constructed high-level semantic feature pyramid network structure at different scales through upsampling and lateral connection. PAN uses bottom-up downsampling to compensate for and enhance positioning information.

The Output section of YOLOv5s uses CIOU_Loss as the loss function, uses non-maximum suppression to screen out excess target frames, predicts image features, and finds the best detection location.

3. Improved YOLOv5s model

YOLOv5s performs well as a general-purpose detection model for datasets in natural scenarios such as COCO and VOC, but the model is less applicable when applied directly to biscuit defect detection in industrial scenarios. Due to its large number of network parameters it can occupy a large amount of memory and lead to a waste of computational resources.

3.1. Model lightweight improvements

3.1.1. Optimize network parameters

In industrial application scenarios, computing resources are very limited, and a large number of

ISSN 2616-5775 Vol. 6, Issue 3: 92-101, DOI: 10.25236/AJCIS.2023.060312

convolution and deep neural networks in the deep learning object detection model cause excessive redundancy, which requires a lightweight design of the original network model. In the model lightweight improvement method, tuning the original network parameters is an efficient lightweight operation method [14]. As a general object detection model, YOLOv5s has too many C3 numbers in the backbone network, which not only increases the amount of calculation and parameters of the algorithm, reduces the inference speed of the algorithm, but also easily leads to the loss of some feature information, thereby increasing the difficulty of biscuit defects detection. In view of the above problems, this paper adjusts the C3 number of YOLOv5s from [3,6,9,3] to [3,3,6,3] to reduce the number of model parameters and prevent the loss of feature information.

3.1.2. Depth separable convolution

Lightweight network Mobilenets, which uses deep separable convolution instead of standard convolution as the basic unit to extract features, under the same convolution [15],reduces computational cost and number of parameters The depth separable convolution consists of two parts including depthWise (DW) and pointwise (PW), as shown in Figure 2.

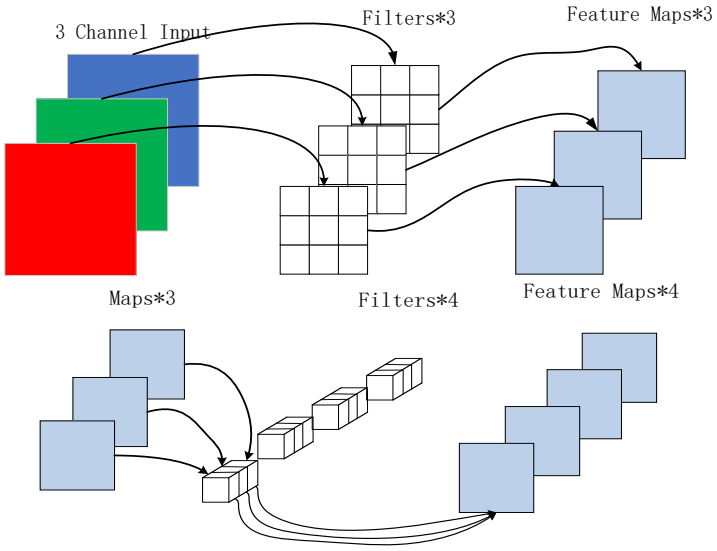


Figure 2: Depthwise convolution (a) and pointwise convolution (b)

Assuming that the input feature map size is $W_{in} \times H_{in} \times M$, the output feature map size is $W_{out} \times H_{out} \times N$, the convolution kernel size is $D_k \times D_k$,then the parameter quantity of the standard convolution is $D_k \times D_k \times M \times N$, the amount of deep separable convolution parameters is $D_k \times D_k \times M + M \times N$, which is $\frac{1}{N} + \frac{1}{D_k \times D_k}$ of the amount of parameter standard convolution.

YOLOv5s is aimed at biscuits defect detection which is single class detection task, and the model is more complex, and overfitting problems are easy to occur during the training process. Therefore, this paper replaces standard convolution with deeply separable convolution to achieve lightweight and balanced accuracy of the network. The specific structure is shown in Figure 4.

3.2. SE attention mechanism

In order to enhance the feature extraction performance of the model for biscuit defects, then realize high accuracy of biscuit defect detection. In this paper, channel attention is used to recalibrate the depth feature map extracted by CNN. At present, SE[11], CBAM[12], CA[13]is the mainstream attention mechanism, of which SE is the most typical channel attention mechanism module, compared with CBAM and CA, SE floating-point number calculation and parameter amount is less, more computation saving, SE attention module can promote the efficient expression of effective features between channels, improve the performance of the model,and the effectiveness and stability of the model has been widely verified.

ISSN 2616-5775 Vol. 6, Issue 3: 92-101, DOI: 10.25236/AJCIS.2023.060312

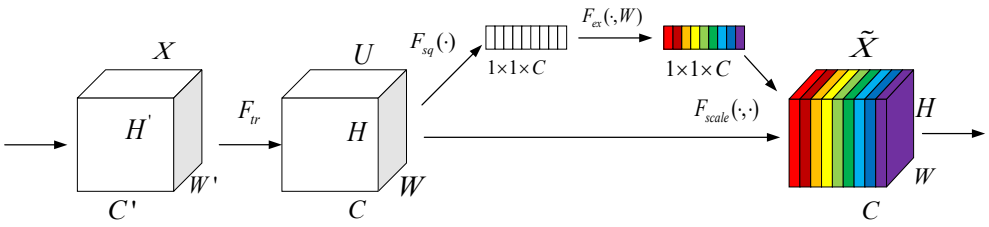


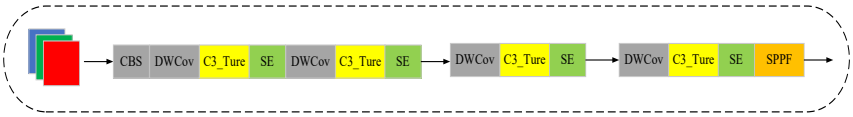
Figure 3: SE module structure

In Figure 3, X represents the input feature map, F_{tr} is a convolution operator, and U is the feature map after convolution operation. As the input of the attention mechanism module, the feature map U compresses the feature layer by the Squeeze operation ($F_{sq}(\cdot)$) named the global average pooling, obtains an output of 1×1×C. After that, the Excitation operation ($F_{ex}(\cdot,W)$) is performed to obtain the weight matrix, the specific formula is as follows:

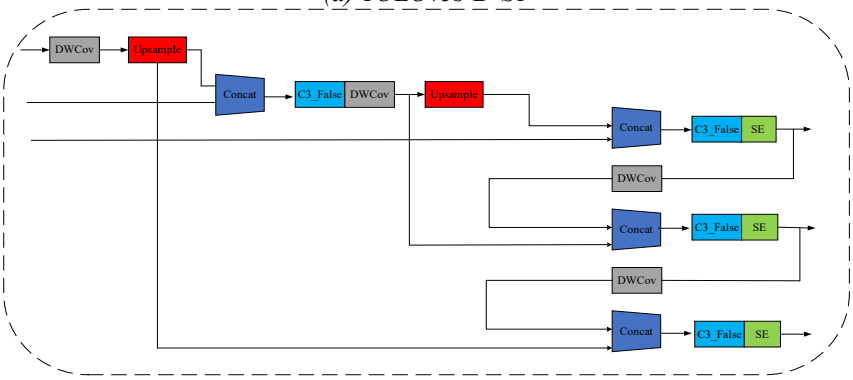
$$s = F_{ex}(z,W) = \sigma(g(z,W)) = \sigma(g(W_2\delta(W_1 z))) \quad (1)$$

where s represents the output after Excitation, z is the output after the Squeeze operation, σ and b are the Sigmoid activation function and the ReLU activation function, respectively. $W_1 \in \mathbb{R}^{\frac{C}{r}\times C}$ and $W_2 \in \mathbb{R}^{\frac{C}{r}\times C}$, are dimensionality reduction operations and ascending operations, respectively, and r represents the proportion of ascending dimension or dimensionality reduction. Finally, the scale operation ($F_{scale}(\cdot,\cdot)$) is performed to multiply the obtained weight matrix with U.

At present, there is no theoretical study to conclude the correlation between the specific embedding location of the SE module and the network performance. Therefore, on the basis of improving the deep separable convolution, this paper designs two structures for the network embedding SE module, as shown in Figure 4. One is to introduce SE modules after the original C3 module of the Backbone layer to generate a new model name YOLOv5s-D-S1, and the other is to embed the SE module after the original C3 module of the Neck layer, named YOLOv5s-D-S2.



(a) YOLOv5s-D-S1



(b) YOLOv5s-D-S2

Figure 4: Improved YOLOv5s network structure

3.3. Loss function improvements

YOLOv5 uses CIOU Loss as the loss function of the Bounding box. Compared to standard IOUs, CIOU Loss CIOU loss increases the detection scale and the loss of length and width, improves the overlap of the detection frame and the real box, and makes the detection box with different overlapping methods have better effects in training, and the CIOU calculation formula is as follows:

$$CIoU = IoU - \frac{\rho^2(b,b^{gt})}{c^2} - \alpha v \quad (2)$$

ISSN 2616-5775 Vol. 6, Issue 3: 92-101, DOI: 10.25236/AJCIS.2023.060312

Among them, IoU ***represents the intersection ratio of the prediction box and the real box,*** (b,b^{gt}) represents the center point between the prediction box and the real box, ρ represents the Euclidean distance, c represents the diagonal distance between the prediction box and the smallest envelope closed box of the real box, α represents the positive equilibrium parameter, and v represents the consistency of the aspect ratio of the prediction box to the real box.

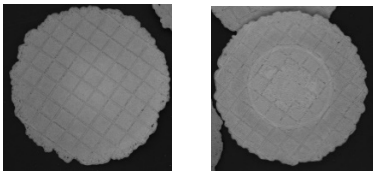
Although $CI\mathrm{oU}$ comprehensively considers the overlapping area, scale and length-width ratio, solves the optimization problem of horizontal and vertical directions between the prediction box and the real box, but the method does not consider the balance problem of difficult and easy samples, and the ambiguity that length-width ratio is relative value. To solve the above problems, this paper uses EIoU loss as the loss function [14], and the formula is as follows:

$$EIoU = IoU - \frac{\rho^2(b,b^{gt})}{c^2} - \frac{\rho^2(w,w^{gt})}{c_w^2} - \frac{\rho^2(h,h^{gt})}{c_h^2} \quad (3)$$

4. Experimental data

4.1. Data acquisition

In this paper, biscuits were used as the research object for recognition experiments, and the image data was collected from a factory in Foshan, Guangdong. The Hikvision MV-CA050-10GM industrial camera was used to shoot biscuits on the production line, and the number of pictures was expanded to 1908 by using random inversion, mirroring and other data augmentation methods, with a target number of 5556, a picture resolution of 1224×1024 pixels, and a JPEG image format. The dataset consists of two categories, one for qualified biscuits, as shown in Figure 5(a), and one for defective biscuits with white rings, as shown in Figure 5(b)

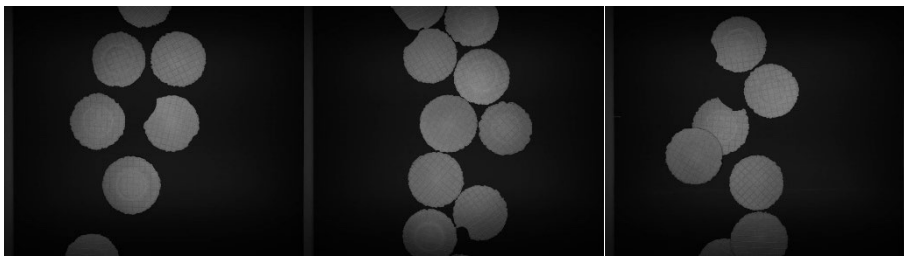


(a) qualified product (b) white ring

Figure 5: Biscuits data set

As can be seen from Figure 5, the surface texture of the good product is clear and free of impurities. The white ring of the defective biscuit is always located in the center of the biscuit image, showing a "double ring" effect under lighting, and the luminosity of white ring differ from one another.

Considering the situation that biscuits are randomly distributed in different positions of the assembly line in actual production, three situations of biscuit sparseness, biscuit adhesion and biscuit stacking were collected respectively, as shown in Figure 6.



(a) Sparse placement diagram (b) Adhesion placement diagram (c) Stacking diagram

Figure 6: Biscuit distribution

4.2. Data labeling

Before training the YOLOv5 model, it is necessary to classify and label the dataset, and use LabelImg labeling software to manually label the biscuits in the picture one by one, with a total of 2 types of labeling, "normal biscuits", "white ring biscuits" that are half-baked in the production process due to temperature and other reasons, labeled as qualified, white, and the labeling process is shown in Figure 7.

ISSN 2616-5775 Vol. 6, Issue 3: 92-101, DOI: 10.25236/AJCIS.2023.060312

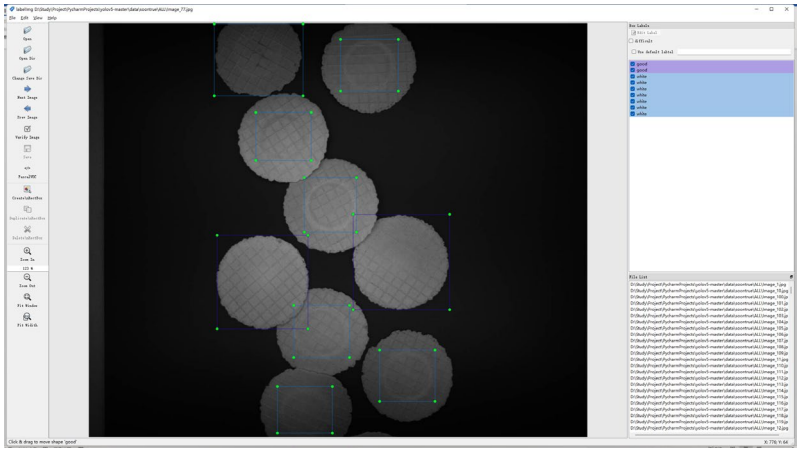


Figure 7: Data annotation

5. Experimental process and results analysis

5.1. Experiment configuration

The training and testing of this experimental model are completed on the Windows 11 operating system, using the Pytorch 1.12 framework and configuring the Nvidia GeForce RTX 3060 graphics card.

5.2. Experiment configuration

In order to objectively evaluate the performance of the improved YOLOv5 model, five evaluation indicators, namely Precision, (P), Recall, (R), mean average precision (mAP), model parameters, and Frames Per Second (FPS), were used to evaluate the network model. P is used to evaluate the accuracy of the model's detection of biscuits; R is used to evaluate the comprehensiveness of model detection. The calculation formula is as follows: equation (1), equation (2):

$$precision = \frac{TP}{FP + TP} \times 100\% \quad (4)$$

$$recall = \frac{TP}{FN + TP} \times 100\% \quad (5)$$

Among them, TP means to detect biscuits and detect the correct quantity; FP indicates the number of biscuits detected but detected incorrectly; FN indicates the number of missed biscuits. AP refers to the average accuracy, which is the area enclosed by the P-R curve and the coordinate axis. mAP is a measure of the accuracy of the network model in each category, and its formula averages the APs of all categories:

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \quad (6)$$

where N is the number of target classes in the dataset, mAP50 is defined as the average AP50 of all classes when the IOU threshold is 0.5, and mAP50 measures the trend of accuracy of the model changing with recall.

5.3. Experimental results and analysis

5.3.1. C3 number comparison experiment

In order to verify the effectiveness of optimizing the C3 number to the model, the number of C3 modules in the backbone network is reduced from [3,6,9,3] to [3,3,6,3] based on YOLOv5s, and the rest is unchanged. The modified algorithm is called Yolov5s-C3, and the results are compared with the original model, and the results are shown in Table 1:

Table 1 shows that by optimizing the number of C3 in the YOLOv5s backbone network, the number of model parameters is reduced by 0.4 M and the detection accuracy is slightly improved, which proves the effectiveness of the improved method.

ISSN 2616-5775 Vol. 6, Issue 3: 92-101, DOI: 10.25236/AJCIS.2023.060312

Table 1: C3 number optimization verification experiment

Algorithm name	P/%	R/%	mAP_{50}/%	Parameter quantity/M	FPS(Frame/s)
YOLOv5s	96.2	96.2	97.9	13.7	66
YOLOv5s-C3	96.1	96.2	98.2(+0.3)	13.3	66

5.3.2. Comparative experiments of deep separable convolution improved

In order to verify the effectiveness of the introduction of deep separable convolution pair model, this paper improves the standard convolution of the backbone layer and neck layer on the basis of YOLOv5s-C3, and the improved algorithm is named YOLOv5-D, as shown in Figure 4, which replaces the standard convolution with deep separable convolution. In this paper, a comparative experiment was used to compare the two algorithms on the biscuit dataset, and the experimental results are shown in Table 2.

Table 2: C3 number optimization verification experiment

Algorithm name	P/%	R/%	mAP_{50}/%	Parameter quantity/M	FPS(Frame/s)
YOLOv5s-C3	96.1	96.2	98.2	13.3	66
YOLOv5s-D	97.3	94.4	97.1	8.7	72

It can be seen from Table 3 that the introduction of deep separable convolution reduces the amount of model parameters by 34.6% and increases the detection speed by 9%, effectively reducing the amount of model parameters and detection speed, but has a great impact on the detection accuracy of the model, and the simple use of depth separable convolution does not meet the use requirements. In order to improve the accuracy of the model, the attention mechanism is introduced, the focus on the detection target is enhanced, the feature extraction ability of the model is improved, and the invalid features of complex background are suppressed.

5.3.3. Attention module comparison experiment

In order to verify the effectiveness of the SE module on the biscuit target detection algorithm and the optimal embedding position of the module, this paper compares the two embedding methods on the basis of Section 3.2. As shown in Figure 3.1, where YOLOv5-D-S1 introduces SE modules after the C3 module of the Backbone layer, and YOLOv5-D-S2 is embedded with the SE module after the C3 module of the Neck layer, and the experimental results are shown in Table 3.

Table 3: Attention Module Validation Experiment

Algorithm name	P/%	R/%	mAP_{50}/%	Parameter quantity/M	FPS(Frame/s)
YOLOv5s -D	97.3	94.4	97.1	8.7	72
YOLOV5s-D-S_1	97.7	95.4	98.5(+1.4)	9.2	70
YOLOV5s-D-S_2	97.4	95.3	97.8(+0.7)	9.2	68

In Table 3, it can be seen from the comparation of Yolov5-D, YOLOV5-D-S_1 and YOLOV5-D-S_2 that, the model detection accuracy is improved by the introduction of SE module, among which the mAP of YOLOV5-D-S1 is increased by 1.4%, and the mAP of YOLOV5-D-S2 is increased by 0.7%. The experimental results show that the introduction of SE module can improve the detection accuracy of the algorithm and make the network pay more attention to the characteristics of biscuit defects. By comparing the results of the two embedding methods, it is found that the network detection accuracy of YOLOV5-D-S1 is improved with less parameter change, because adding the attention mechanism to the feature extraction layer allows the model to learn more biscuit defect features. Therefore, the first SE embedding method is selected for the improved algorithm in this paper. In order to further verify the effectiveness of the SE module, CBAM, ECA, CA attention are introduced on the basis of Yolov5-D, compared horizontally with the first SE embedding method, and the experimental results are shown in Table 4.

Table 4: Attention module transverse contrast experiment

Algorithm name	P/%	R/%	mAP_{50}/%	Parameter quantity/M	FPS(Frame/s)
YOLOv5s -D	97.3	94.4	97.1	8.7	72
YOLOv5s -D-CA	96.1	96.2	98.2(+0.3)	9.8	65
YOLOv5s -D-ECA	84.2	89.8	97.9(+0.8)	9.7	66
YOLOv5s -D-CBAM	95.5	94.7	98.1(+1.0)	9.8	62
YOLOv5s -D-SE	97.7	95.4	98.5(+1.4)	9.2	68

It can be seen from the horizontal comparison experiment of the attention module in Table 5 that, by embedding attention module, the biscuit detection accuracy was improved, among which, CA, ECA and CBAM increased the detection accuracy by 0.5%, 0.8% and 1%, respectively, while the detection accuracy of SE increased the most, and the detection accuracy reached 98.5%. Comparing the number of

ISSN 2616-5775 Vol. 6, Issue 3: 92-101, DOI: 10.25236/AJCIS.2023.060312

parameter, after introducing attention module, the number of parameter increased to a certain extent, but the number of model parameters introduced SE was the least. In summary, the introduction of SE module after the C3 module of the Backbone layer proposed in this paper improves the detection performance of biscuits better than other attention modules

5.3.4. Ablation experiments

To verify the effectiveness of the improved algorithm proposed in this paper, ablation experiments were performed on each module. Firstly, the number of C3 in the backbone network is optimized on the basis of the original algorithm YOLOv5s. Second, use deep separable convolution instead of partial standard convolution. Then, SE modules are introduced after the C3 module in the Backbone layer. Finally, EIOU Loss is used as the loss function to generate the final improved algorithm model, which is compared with the original algorithm YOLOv5s. P, R, mAP50, parameter quantity as evaluation index. The results of the comparative experiments are shown in Table 5.

Table 5: Attention module transverse contrast experiment

Method	C3	DW	SE	EIoU	P/%	R/%	mAP_{50}/%	Parameter quantity/M	FPS (Frame/s)
YOLOv5s					96.2	96.2	97.9	13.7	66
Module1	√				96.1	96.2	98.2	13.3	66
Module2	√	√			97.3	94.4	97.1	8.7	72
Module3	√	√	√		97.7	95.4	98.5	9.2	68
Module4	√	√	√	√	97.8	95.8	99.2	9.2	67

It can be seen from the experimental results in Table 5 that compared with the original YOLOv5s, the number of C3 in the YOLOv5s backbone network is reduced from [3,6,9,3] to [3,3,6,3], which can not only reduce the number of parameters, but also slightly improve the detection accuracy of biscuit defects. Model 2 uses deep separable convolution instead of standard convolution on the basis of model 1, which reduces the number of parameters by 31.6%, but has a great impact on the detection accuracy. The SE module is introduced on the basis of model 2 to make the useful feature information more prominent, so as to improve the detection accuracy of biscuits, although the method increases some parameters, but the mAP of the network model increases by 1.4%; Model 4 uses EIoU loss as the regression loss function to form the final improved model, compared with model 3, the number of parameters did not increase, and the mAP of the model reached 99.2%, which was 1.3% higher than the original model, which proves the superiority of the EIoU loss function to improve the performance of the model. The above experimental results prove the effectiveness and superiority of the improved algorithm proposed in this paper in biscuit detection.

Figure 8 shows the mAP comparison between the improved algorithm and the original algorithm during training, from which it can be clearly seen that the detection accuracy of the final improved model is better than that of the original YOLOv5s, and it is greatly improved, and better detection results are achieved.

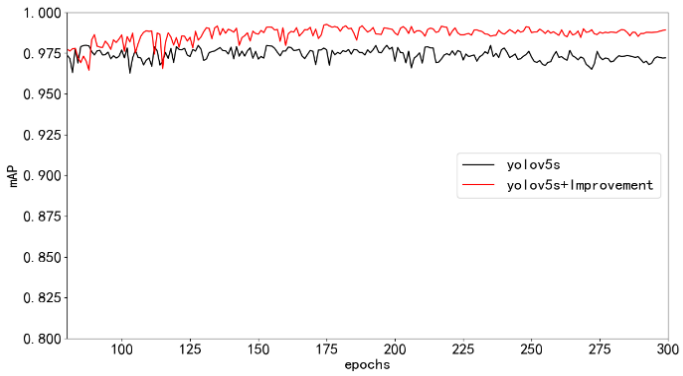
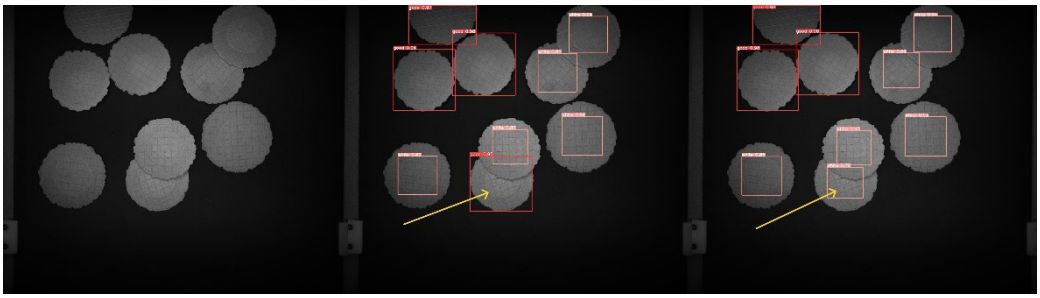


Figure 8: Comparison of mAP between the improved algorithm and the original algorithm during training

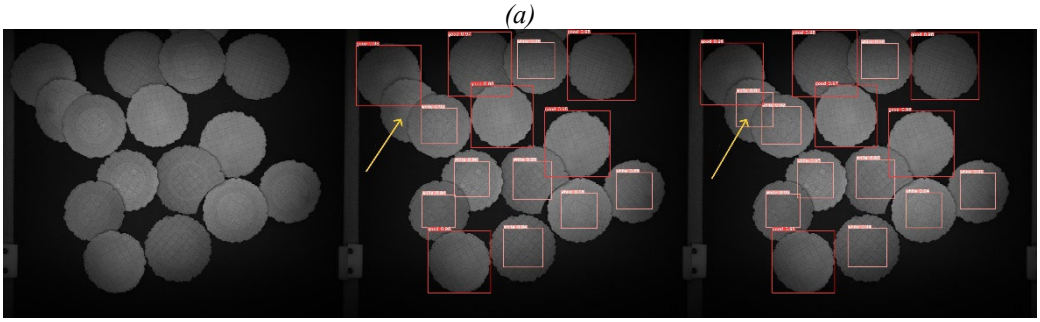
Figure 9 shows the test results of the original figure, YOLOv5s, and improved YOLOv5s from left to right. The experimental results show that YOLOv5s have false detection and missed detection when the defects of biscuits in groups (a) and (b) are occluded, while the improved YOLOv5s can accurately detect biscuit defects. In group (c) biscuits, the defects were similar to the biscuit pattern, the

ISSN 2616-5775 Vol. 6, Issue 3: 92-101, DOI: 10.25236/AJCIS.2023.060312

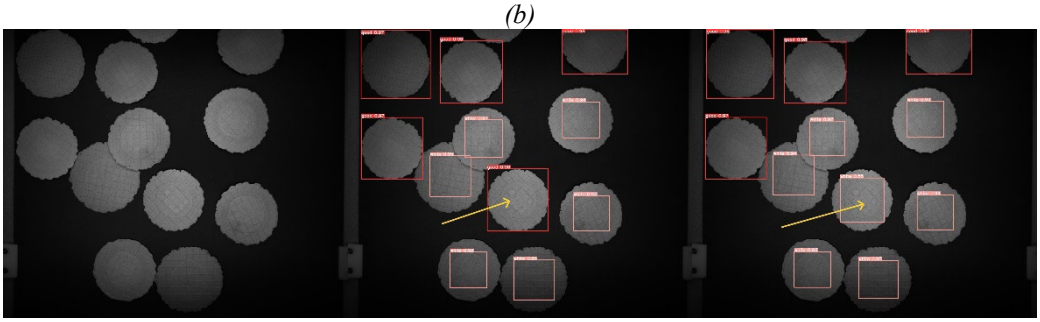
characteristics were not obvious, and YOLOv5s had false detections, but the improved YOLOv5s could accurately predict the defects. The recognition results show that the recognition effect of the improved network model for some targets with missing feature information and similar features is significantly improved, which further proves the effectiveness of the improved network.



(a)



(b)



(c)

Figure 8: Model before and after improvement

5.3.5. Ablation experiments

To further validate the performance of the method in this paper, a comparison test is set up to compare the algorithm in this paper with the mainstream algorithms in the field of target detection under the same experimental environment. Faster R-CNN, a representative of two-stage target detection algorithms, and SSD, YOLOv3[15], YOLOv4-tiny[16], and YOLOv5, representatives of one-stage target detection algorithms, were selected. The experimental results are shown in Table 6.

Table 6: Comparative experiment with the current mainstream methods

Algorithm name	P/%	R/%	mAP_{50}/%	Parameter quantity/M	FPS(Frame/s)
SSD	85.9	88.9	91.6	91.9	76
YOLOv4-ting	96.6	96.3	97.7	22.4	219
FasterR-CNN	96.2	95.6	98.1	108	11
YOLOv5s	96.2	96.2	97.9	13.7	66
Algorithm of this paper	97.9	95.8	99.2	9.2	67

It can be seen from Table 6 that compared with other mainstream detection models, the improved algorithm proposed in this paper has the least number of model parameters and the highest model accuracy. Compared with the original network model, the accuracy is improved by 1.3%, the number of model parameters is reduced by 32.8%, and the detection speed can reach 67 frames per second. Compared with the two-stage algorithm Faster R-CNN, the improved algorithm has obvious parameter quantity and speed advantages, and compared with SSD and YOLOv4-ting, it also has improved detection accuracy. In summary, the improved algorithm proposed in this paper has the highest detection accuracy, better detection speed, best lightweight degree in biscuit defect detection, and the overall

ISSN 2616-5775 Vol. 6, Issue 3: 92-101, DOI: 10.25236/AJCIS.2023.060312

performance of the model is more prominent.

6. Conclusions

A new biscuit defect detection algorithm based on deep convolutional neural network is designed for the problem of defect leakage and misdetection of biscuits. By optimizing the number of C3 modules in the network and introducing the deep separable convolution in MobileNets network for feature extraction, the number of model parameters is reduced and the detection speed is improved; the SE attention module is embedded in the feature extraction layer and EIOU is used as the loss function to improve the model accuracy. The experimental results show that the average single-image detection time of the improved network can reach 15ms and the mAP reaches 99.2%. The problem of biscuit defect leakage and false detection is solved, and it has obvious advantages when compared with other mainstream models.

References

[1] Cheng JF, Fang GWS, Gao HF. Research progress of machine vision technology for surface defect detection [J/OL]. Application Research of Computers. 1-13[2023-02-26]. https: // doi. org/ 10. 19734/ j. issn. 1001-3695.2022.08.0426.

[2] Ren Shaoqing, He Kaiming, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2017, 39 (6): 1137-1149.

[3] Liu Wei, Anguelov D, Erhan D, et al. SSD: single shot multibox detector [C]// European Conference on Computer Vision, [s.L.]: Springer, Cham, 2016: 21-37.

[4] Redmon J, Farhadi A. YOLO9000: better, faster, stronger [J]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 6517-6525.

[5] Wan Guang et al. Ceramic tile surface defect detection based on deep learning [J]. Ceramics International, 2022, 48(8): 11085-11093.

[6] Ma Zhuxi, Li Yibo, Huang Minghui, et al. A lightweight detector based on attention mechanism for aluminum strip surface defect detection [J]. Computers in Industry, 2022, 136: 103585.

[7] Su Z, Han K, Song W, et al. Railway fastener defect detection based on improved YOLOv5 algorithm[C]// 2022 IEEE 6th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). 0.

[8] Zheng Liaomo, Wang Xiaojie, Wang Qi, et al. A fabric defect detection method based on improved YOLOv5 [C]// 2021 the 7th International Conference on Computer and Communications (ICCC), Piscataway, NJ: IEEE, 2021: 620-624.

[9] Ultralytics. YOLOv5 [EB/OL]. [2022-4-10]. https:// github.co m/ultralytics/yolov5.

[10] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications [J]. International Journal of Computer Vision, 2017, 5(8): 122-131.

[11] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[12] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[13] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C] // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13713-13722.

[14] ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression [J]. Neurocomputing, 2022, 506: 146-157.

[15] Redmon J, Farhadi A. YOLOv3: an incremental improvement [EB/OL] [2022-5-25]. https://arxiv. org/pdf/1804.02767.pdf.

[16] Jiang X, Hu H, Liu X, et al. A smoking behavior detection method based on the YOLOv5 network[J]. 2022.