

Received 18 June 2022, accepted 10 July 2022, date of publication 25 July 2022, date of current version 4 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3193775

RESEARCH ARTICLE

Defect Identification of Adhesive Structure Based on DCGAN and YOLOv5

YONG JIN¹, HUIFANG GAO¹, XIAOLIANG FAN², HASSAN KHAN¹, AND YOUXING CHEN¹

¹School of Information and Communication Engineering, North University of China, Taiyuan 030051, China

²School of Earth Sciences and Engineering, Nanjing University, Nanjing 210023, China

Corresponding author: Yong Jin (xiandaijiance601@163.com)

This work was supported by the Shanxi Province Natural Science Foundation under Grant 201901D111155.

ABSTRACT To overcome the problem of fewer sample and uneven distribution of defect type in defect detection of adhesive structure parts, a defect identification approach based on DCGAN and YOLOv5 is proposed. The above problems are solved by fine-tuning the structure and loss function of DCGAN, the generated high-quality defect images and the extended defect dataset are the basis for accurate identification with YOLOv5. The EIOU loss function is utilized in the YOLOv5 network, the mAP value and recall increase by 3.9% and 10.5% compared with the GIOU loss function, but the precision decreases. To solve this problem, the feature extraction capability of the network is enhanced by incorporating the CBAM after the C3 module in the YOLOv5 network. The mAP, precision, and recall of the optimized YOLOv5 algorithm are improved to 78.6%, 77.2%, and 76%, respectively, the precision compared to the original model improved by 10.6%. The results demonstrate that the improved YOLOv5 model can effectively identify defects of adhesive structure.

INDEX TERMS CBAM, DCGAN, defect identification, EIOU, YOLOv5.

I. INTRODUCTION

Due to the technological limitations and complexity of the environment, there are many defects in the production of adhesive structure, such as debonding, cracking, and delamination. The ability to appropriately identify these defects is critical for optimizing production techniques and improving quality.

At present, the adhesive structure defect detection method based on X-ray imaging, which is completed by manual participation in defect types identification, is not only difficult to ensure the accuracy of judgment due to certain subjectivity but also time-consuming and labor-intensive. In recent years, additional image defect type identification methods have been adopted, including Threshold segmentation [1], Support Vector Machine (SVM) [2]–[4], and Artificial Neural Network (ANN) [5]. However, the above method is difficult to be applied to the recognition of multi-type defect feature images with no obvious difference in grey level, and the location information of defects cannot be detected. In addition,

The associate editor coordinating the review of this manuscript and approving it for publication was Chuan Li.

since adhesive structure defect images may contain multiple defects, each image cannot be simply classified into a certain category for those images with two or more defects simultaneously, but a target detection algorithm similar to the YOLO network should be used to find all the defects in each image and give the location information of the detect area. Compared with previous algorithms, YOLOv5 is faster and lighter, which to some extent is the best performing algorithm in the YOLO family. Although the YOLO network has fast a detection speed, the detection effect for small targets is poor. Multiple scholars [6]–[8] have developed an improved YOLOv5 network, which promotes the recall rate, accuracy, and mAP. Therefore, the defect identification method based on the YOLOv5 network can be used as an effective method to identify the defect of adhesive structure. However, there is still room for further optimization structure and adjustment parameters of the YOLOv5 model to achieve specialized and efficient detection of small targets such as adhesive structure defects.

Due to the improvement of the production process, the increase in yield and the decrease in defective sample images emerge. The dataset established on this basis will lead to

overfitting of the deep learning network, resulting in the inability to achieve efficient defect detection. However, data enhancement methods can expand the dataset and solve the problem of small sample quantity. The in-depth research has brought forth an endless variety of image data enhancement techniques, which can be mainly divided into two categories: non-generative and generative data enhancement methods [9]. In detail, non-generative data enhancement methods, such as rotation and color enhancement, only expand the number of data sets, and cannot greatly improve the generalization ability of the network model. Generative data enhancement methods, such as Deep Convolutional Generative Adversarial Network (DCGAN) [10]–[12], can not only increase volume of dataset, but also increase the diversity of images, thereby improving the generalization ability of the trained model. Based on the above, the improved DCGAN network can be used to increase the dataset in view of the problem of fewer defect samples and the unbalanced distribution of adhesive structure. On the one hand, the problem of small number of samples may be overcome; on the other hand, the problem of overfitting can be easily solved in the network due to the single kind of dataset.

Based on the above background, this study proposes an adhesive structure defect recognition method based on DCGAN and YOLOv5, which the improved DCGAN network expanding the dataset and the optimized YOLOv5 network training and testing the expanded defect dataset. The following are the primary contributions of our work:

(1) In order to obtain high-quality generated images, the structure of DCGAN network has been upgraded, including the network layer, convolution kernel parameters, activation function, and loss function.

(2) Optimize YOLOv5's structure as follows: the CBAM mechanism is introduced after the C3 module to fuse the features. The YOLOv5 network's GIOU loss function is replaced with the EIOU loss function. As a result, the position information of the defect can still be obtained efficiently and the convergence be accelerated.

II. GUIDELINES FOR MANUSCRIPT PREPARATION

A. IMPROVE THE DCGAN NETWORK

DCGAN is a deep convolutional generative adversarial network, which is composed of generative network and a discriminant network [13]. Compared with the GAN model, DCGAN optimizes network structure using the concept of a deep convolution network, to improve the quality of sample generation and convergence speed. Specific improvement measures include: in the generative network, using transpose convolution instead of pooling layer to achieve up-sampling, using BN to solve the training problem caused by poor initialization, and using Tanh activation function in the output layer and ReLU activation function in other layers to reduce the risk of vanishing gradient. In the discriminant network, with step-size convolution instead of pooling layer for down-sampling, using BN to stable training, using sigmoid function

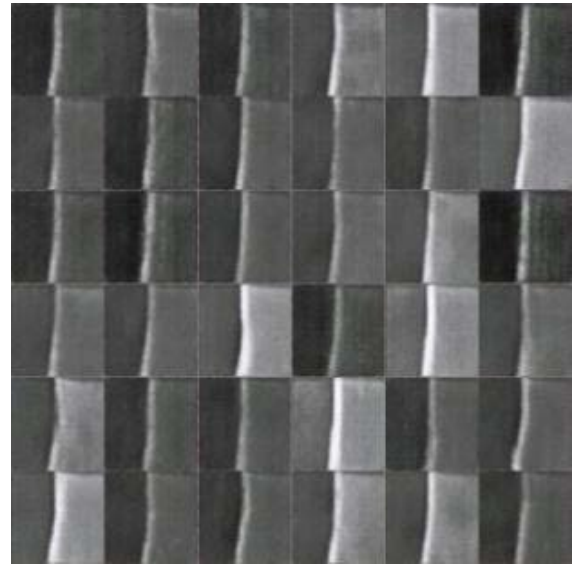


FIGURE 1. Generated image for 4500 iterations.

for output layer and LeakyReLU activation function for other layers to alleviate the problem of gradient loss.

Nevertheless, there are still some images with poor quality that are difficult to distinguish in the generated samples when DCGAN generates new defect images, as shown in Fig.1. These images are seriously blurred and blurry. Using these images for network training not only reduces the generalization ability of the network, but also reduces the accuracy of target detection. Therefore, DCGAN must be optimized to obtain high-quality generated images.

Based on DCGAN, the network layers of generative network and discriminant network are added, and the output image resolution is increased to 256×256 . In order to address the problem of parameter oscillation, it is proposed that each layer of the convolutional network be joined by GN (GroupNorm). The transpose convolution kernel in the generative network is set as 4×4 convolution with a step size of 2, and the number of convolution kernels is set as [512, 512, 512, 256, 128, 64] to improve feature extraction and reduce calculation difficulties. Taking Block5 of the generative network as an example, DCGAN performs convolution operation of input features with kernel size of 5×5 and step size of 2, and the number of parameters is $5 \times 5 \times 128 \times 64 = 204800$, whereas this paper is $4 \times 4 \times 128 \times 64 = 131072$. It can be seen that the parameters of the convolution module designed in this paper have been considerably lowered, which aids in improving the calculation efficiency. ReLU activation function is utilized in all layers except Tanh activation function in the output layer on the basis of BN normalization to overcome the problem of gradient disappearance and accelerate the model convergence. The discriminant network's convolution operation corresponds to the generative network's transpose convolution operation one by one, and all but the last layer

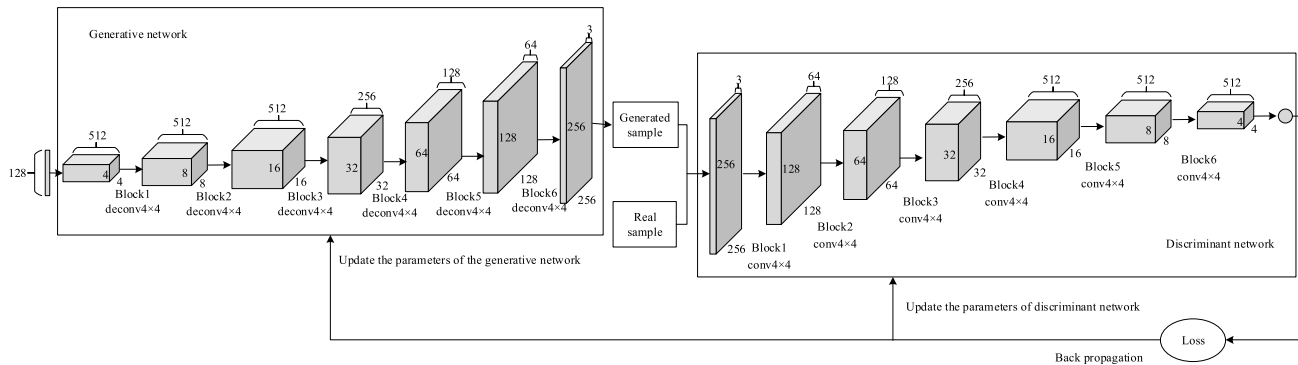


FIGURE 2. The structure of DCGAN.

employ the LeakyReLU activation function based on GN normalization. The specific network is shown in Fig.2.

B. IMPROVED DCGAN LOSS FUNCTION

Real data x obeys $P_{data(x)}$, and noise z obeys noise distribution $P_{z(z)}$. Formula (1) shows the loss function of the DCGAN generative network, whereas Formula (2) shows the loss function of the discriminant network, which contains the discrimination error of the real and generated image [14]:

$$Loss_G = E_{z \sim P_{z(z)}} [\log (1 - D(G(z)))] \tag{1}$$

$$Loss_D = E_{x \sim P_{data(x)}} [\log D(x)] + E_{z \sim P_{z(z)}} [\log (1 - D(G(z)))] \tag{2}$$

where G is the generative network, D is the discriminant network, $G(z)$ is the image obtained by the generator, $D(G(z))$ is the discrimination probability of the generated image, and $D(x)$ is the discrimination probability of the real image [15].

The DCGAN network solution is a process of maximizing the discriminant network and minimizing the generative network. The optimization objective of the training process is shown in formula (3):

$$\min_G \max_D V(D, G) = E_{x \sim P_{data(x)}} [\log D(x)] + E_{z \sim P_{z(z)}} [\log (1 - D(G(z)))] \tag{3}$$

where $V(D, G)$ is cross entropy loss. The purpose of the generative network G is to make the generated image as close to the real image as possible, that is, $D(G(z))$ should reach the maximum value as possible, at which point $V(D, G)$ should reach the minimum value, corresponding to \min_G of the formula; The purpose of the discriminant network D is to judge the authenticity of the input image more accurately, that is, $D(x)$ should be as large as possible and $D(G(z))$ should be as small as possible, corresponding to of the formula.

The training discriminant network calculates the JS distance between the generated and real sample distribution, whereas the training generator minimizes the JS distance [16]. JS divergence becomes constant when there is no overlap between the two samples. At this point, the generator’s loss function becomes constant, causing the gradient of the

generator to vanish and the network parameters cannot be updated. The improved method is to employ the cross entropy loss function with gradient penalty term, perform random number interpolation \hat{x} between real and generated data, and construct gradient penalty term GP using the weight coefficient λ and interpolation \hat{x} , as shown in formula (4):

$$Loss_{GP} = E_{\hat{x} \sim P_{\hat{x}}} [\| \nabla_{\hat{x}} D(\hat{x}) \|_2 - 1]^2 \tag{4}$$

where $P_{\hat{x}}$ represents the random sample data distribution in the generated sample set region, the real sample set region, and the intermediate region. $\hat{x} = \alpha x + (1 - \alpha) z$, $\alpha \sim \mu(0, 1)$. x follows the real data distribution and z follows the generated data distribution.

At this point, the discriminant network loss function can be expressed as:

$$L_D^{GP} = Loss_D + \lambda Loss_{GP} \tag{5}$$

where λ is the coefficient of gradient penalty term, and the value of λ in this paper is 10.

C. DCGAN IMAGE GENERATION EXPERIMENT

1) EXPERIMENTAL DESCRIPTION

The purpose of this experiment is to verify the effectiveness of the improved DCGAN network. This experiment is run in the GPU environment, with Python as the experiment platform’s programming language and the PyTorch framework as its basis. The optimizer of the model is Adam. The hyperparameters are set as follows: The initial learning rate of the model is set to 0.0002. The epoch for training is 450. The batch size of each training is set to sixteen times to improve the training speed. The resolution size of the input image is set at 256×256 pixels according to the common size, and the default parameters are utilized for other parameters.

2) EFFECT ANALYSIS OF DCGAN IMPROVEMENT

Fig.3 shows the images obtained when the number of iterations is 100, 1000, 2500, and 4500 during the training of the improved DCGAN model. At the beginning of training (Fig.3(a)), the generated images are noisy, vague, and without any form. With the training to the 1000th (Fig.3(b)), the

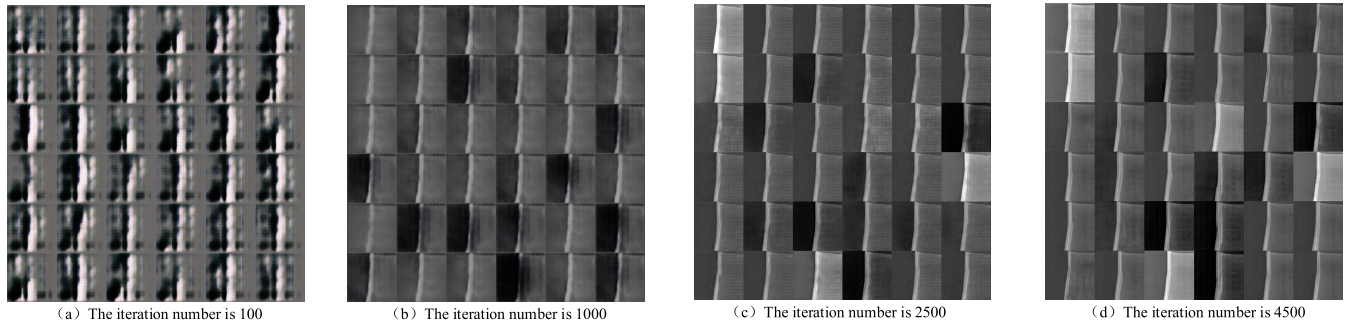


FIGURE 3. Generated images of improved DCGAN.

general outline is already available and the basic structure of the image can be seen, but there are still unclear problems. At the 2500th iteration (Fig.3(c)), the images become clearer, and a clearer result is attained after the 4500th iteration. The observation shows that the general trend of its generation effect becomes clearer and clearer with the increase of iterations.

In addition to the generated images reflecting the effect of improved DCGAN training, the training process of improved DCGAN can also be dynamically observed through the loss function curve. The DCGAN training process not only reduces the loss of the generation network and discriminant network but also balances the process. Tensorboard is used to draw the curves of two networks, as shown in Fig.4. The horizontal coordinate in the diagram represents the iteration times(iter), the vertical coordinate D_loss represents the discriminator loss, g_loss represents the generator loss, gp represents the gradient penalty term, $x_real_d_loss$ represents the loss function obtained from the real image input into the discriminant network, and $x_fake_d_loss$ represents the loss function obtained from the generated image input into the discriminant network. The loss curve of $x_real_d_loss$ and $x_fake_d_loss$ tends to be stable and hovers around 0.5 after 1000 iterations, indicating the discriminator's discriminant ability and the generator's generating ability have reached a certain degree, and the discriminator can no longer distinguish the real image from the generated image. In the early stage of training, both g_loss and d_loss decline steadily, indicating that the two networks have losses that can be optimized for each other, and the generated image quality is also improving consistently. With the increase in the number of iterations, the values of both no longer change significantly, and the model gradually tends to be stable and convergent.

III. IMPROVED YOLOv5s MODEL

YOLO is a target detection algorithm that is based on regression [17]. To save storage cost and improve inference speed, the YOLOv5s model with the minimum network depth and width is adopted as the reference model for adhesive structure defect identification. As illustrated in Fig.5, the YOLOv5s network model is divided into four parts: Input, Backbone, Neck, and Prediction [18]. The input module mainly performs

Mosaic data enhancement, adaptive image scaling, and adaptive anchor box calculation. Mosaic data enhancement uses a combination of four images to enrich data diversity, and the use of adaptive anchor box calculation is beneficial to improve the detection speed, and the preprocessing results are shown in Fig.6. The Backbone includes the structures of Focus, C3, and Spatial Pyramid Pooling(SPP). It is used to extract image features, in which the Focus module carries out sampling operations on the image and stacks the sampled slices to ensure that feature extraction is sufficient. The C3 structures are designed in the YOLOv5 network, with the C3_1 structure in the Backbone and the C3_2 structure in the Neck. The SPP module uses the maximum pooling layer with four convolution kernels of varying sizes ($k = \{1 \times 1, 5 \times 5, 9 \times 9, 13 \times 13\}$) to achieve feature fusion at multiple scales. The Neck consists of Feature Pyramid Network(FPN) and Path Aggregation Network(PAN). FPN up-samples the image from top to bottom, fusing the extracted features with the backbone network, whereas PAN down-samples the image from bottom to top, fusing the extracted features with the FPN. Prediction includes the bounding box loss function and Non Max Suppression(NMS). Feature Maps of three scales are included in the output exports, which are used to detect large, medium, and small targets. The NMS eliminates redundant prediction boxes, and the information of the prediction boxes with the highest confidence is retained to complete the target detection process.

Although YOLOv5s has a good performance in precision and accuracy, it still has a bottleneck when it comes to detecting adhesive structure defects. The following are specific improvement measures:

(1) CBAM attention mechanism

The YOLOv5s model is improved by introducing an attention mechanism to better deal with defect information. The Attention mechanism mainly acts on the feature graph and enhances the feature expression ability of the network. A large number of experiments have proved that using channel attention first and then using spatial attention can achieve the best effect for network learning. The Convolutional Block Attention Module (CBAM) [19], combined the channel and spatial attention mechanism, retaining more useful feature information. Fig.7 depicts the CBAM structure.

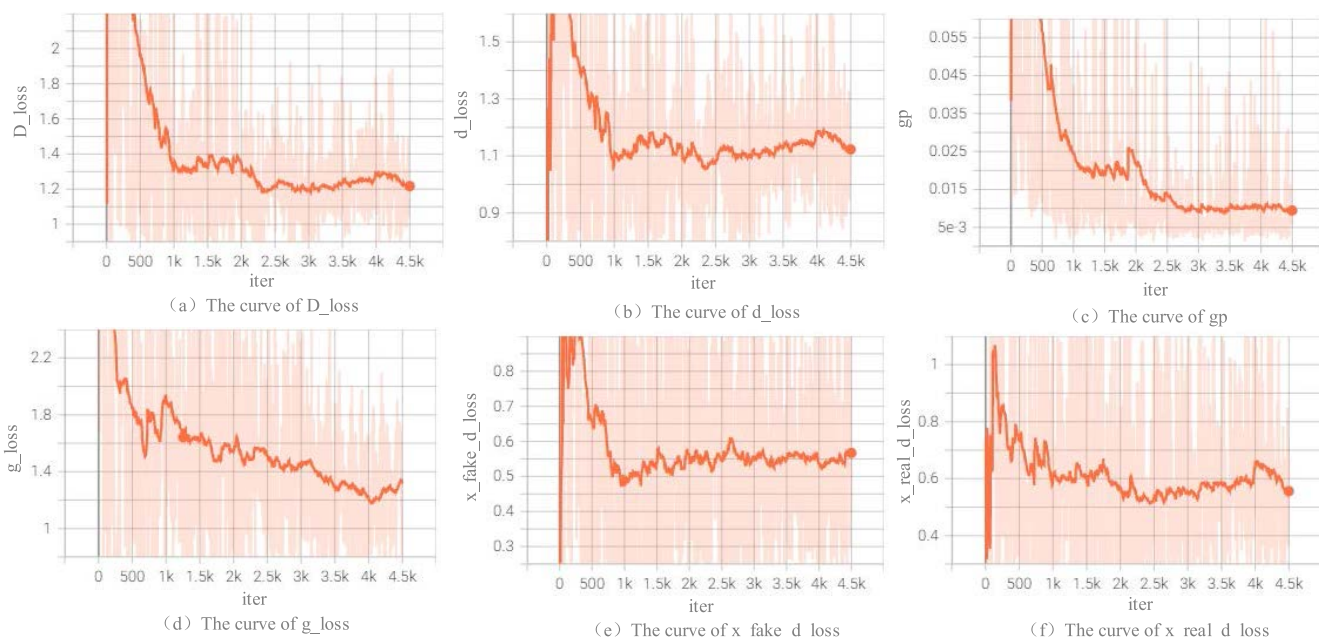


FIGURE 4. Loss function curve.

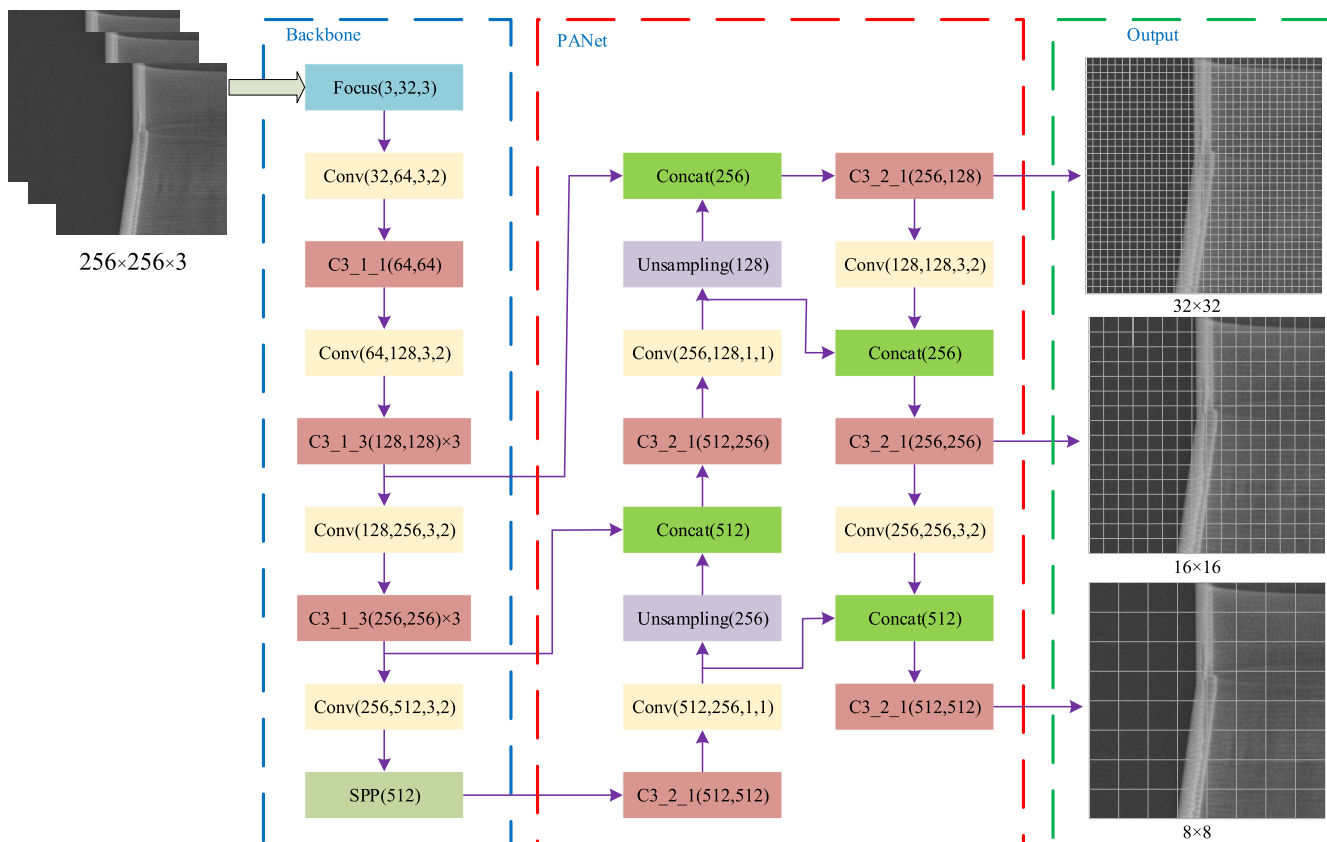


FIGURE 5. Overall architecture of YOLOv5s algorithm.

In CBAM, the input $H \times W \times C$ feature graph F is max-pooled and mean-pooled respectively to produce two $1 \times 1 \times C$ feature graphs, which are then sent to the multi-layer perceptron(MPL). The one-dimensional channel attention graph

$Mc(F)$ is obtained by sum and Sigmoid activation function, and $Mc(F)$ is multiplied by input feature graph F to obtain the channel attention adjusted feature graph $F1$. Then, $F1$ is performed max-pooling and mean-pooling to obtain two

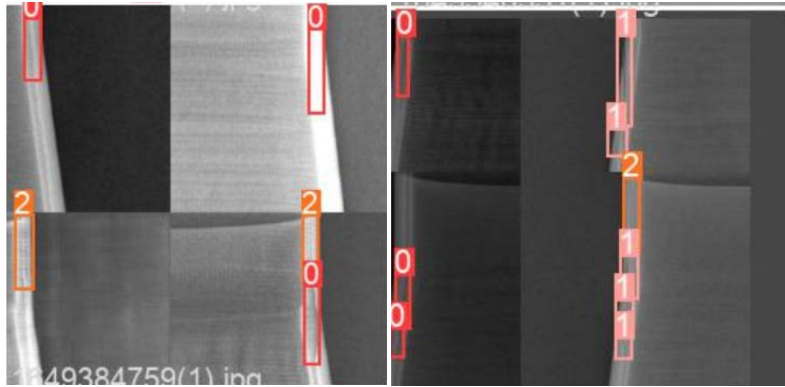


FIGURE 6. Mosaic data enhancement.

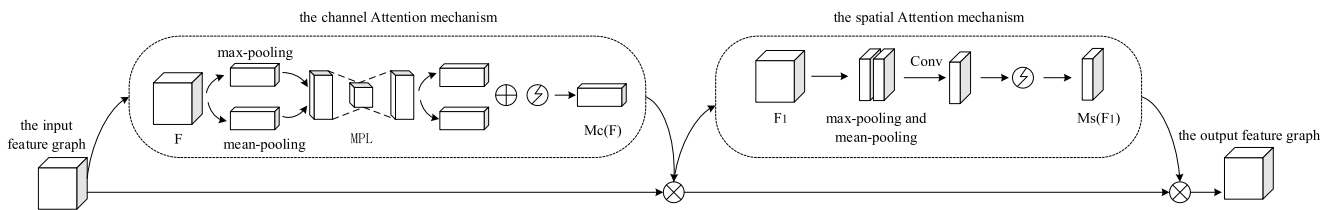


FIGURE 7. The structure of CBAM.

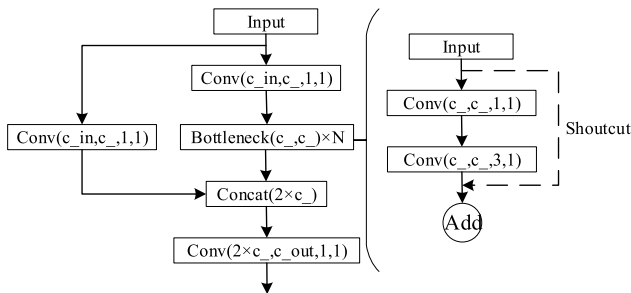


FIGURE 8. The C3 structure.

$H \times W \times 1$ feature graphs, and the two two-dimensional vectors generated after pooling are spliced and convolved to finally generate the two-dimensional spatial attention graph $Ms(F1)$, which is then multiplied with the feature graph $F1$.

C3 is a structure designed based on the ideas of CSP-Net, as shown in Fig.8. C3-1 is located in the Backbone of YOLOv5s, and its residual component is designed based on Resnet. By adding shortcut between convolutional layers, the computation of the network model is reduced and the operation efficiency of the network is accelerated. C3-2 is located in the Neck of YOLOv5s without shortcut. It mainly carries out convolution operation on the input feature map and fuses the extracted feature information.

As shown in Fig.9, the CBAM attention mechanism is added after the C3 module in this research.

(2) Loss function

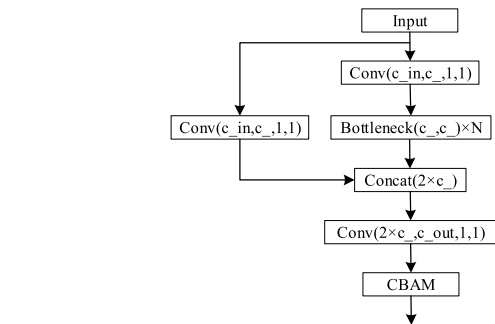


FIGURE 9. The C3 structure of adding attention mechanism.

The development process of the regression loss function in recent years is as follows: $IOU_Loss \rightarrow GIOU_Loss \rightarrow DIOU_Loss \rightarrow CIOU_Loss \rightarrow EIOU_Loss$.

IOU_Loss is shown in formula (6), when the prediction box and the real box do not intersect, that is $A \cap B = 0$, the loss function is no gradient back at this time. Furthermore, when the size of the prediction box and the real box are the same, IOU may be different, as illustrated in Fig.10, and the IOU_Loss function is unable to distinguish between the two cases.

$GIOU_Loss$ is shown in formula (7), if the prediction box and real box are part of a containment relationship, $GIOU$ will still degenerate into IOU . Meanwhile, the $GIOU$ continues to have problems, such as the unstable target box regression and the easy divergence during training.

$DIOU_Loss$, as shown in formula (8), ignores the aspect ratio of the prediction box and the real box, focusing instead

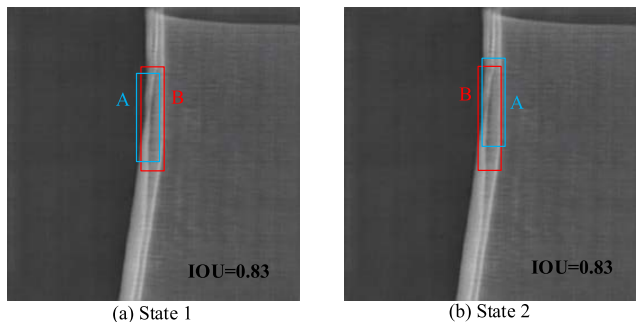


FIGURE 10. The special status of IOU_Loss.

on the overlapping area of the bounding box and the center point distance of b and b^{gt} .

CIOU_Loss adds a penalty term on the basis of IOU_Loss, as shown in formula (9). It takes into account the aspect ratio of the predicted box to fit the target box, resulting in faster network convergence and higher regression positioning accuracy during training.

EIOU_Loss, which is shown in formula (12), is obtained on the basis of CIOU_Loss, and it not only takes into account the central point distance and the aspect ratio, but also the true discrepancies in the target and anchor boxes' widths and heights. The EIOU_Loss function directly minimizes these discrepancies and accelerates model convergence.

$$IOU_Loss = 1 - IOU = 1 - \frac{A \cap B}{A \cup B} \quad (6)$$

$$GIOU_Loss = 1 - GIOU = 1 - \left(IOU - \frac{|C - A \cup B|}{|C|} \right) \quad (7)$$

$$DIOU_Loss = 1 - DIOU = 1 - \left(IOU - \frac{\rho^2(b, b^{gt})}{c^2} \right) \quad (8)$$

$$CIOU_Loss = 1 - CIOU = 1 - \left(IOU - \frac{\rho^2(b, b^{gt})}{c^2} - \alpha\gamma \right) \quad (9)$$

$$\alpha = \frac{\gamma}{1 - IOU + \gamma} \quad (10)$$

$$\gamma = \frac{4}{\pi^2} \left(\arctan \frac{\omega^{gt}}{h^{gt}} - \arctan \frac{\omega}{h} \right)^2 \quad (11)$$

$$EIOU_Loss = 1 - EIOU = 1 - \left(IOU - \frac{\rho^2(b, b^{gt})}{c^2} - \frac{\rho^2(\omega, \omega^{gt})}{c_\omega^2} - \frac{\rho^2(h, h^{gt})}{c_h^2} \right) \quad (12)$$

where A is the prediction box and B is the real box; C is the area of the smallest enclosing rectangle of the real box and the prediction box; b and b^{gt} represents the center points of the prediction and the real boxes, respectively; $\rho^2(\bullet)$ represents the Euclidean distance; c is the diagonal distance of C ; ω^{gt} , h^{gt} and ω , h represent the width and height of the real and prediction boxes, respectively; c_ω and c_h represent

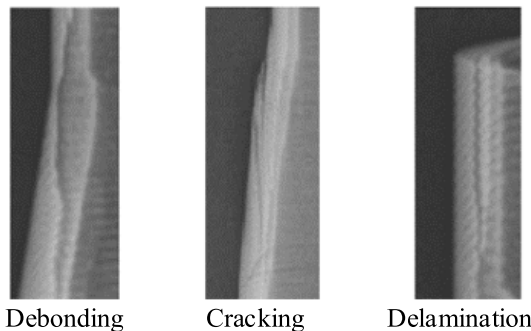


FIGURE 11. Defect image of adhesive structure.

the width and height, respectively, of the smallest enclosing box covering the two boxes.

IV. YOLOv5s DEFECT IDENTIFICATION EXPERIMENT

A. PREPARATION OF ADHESIVE STRUCTURES DEFECT DATASET

The data set used in this experiment comes from the X-ray test samples of multi-layer metal and non-metal bonded tubular specimens in the past two years. The adhesive structure defects of the sample mainly include: debonding, cracking, and delamination, as shown in Fig. 11. Among them, debonding refers to poor bonding between layers, and it can be seen from the figure that there are obvious black images between inner and outer layers. Cracking is the outer or inner layer of cracking defects, the image is black dendritic. Delamination is the outer or inner layers that are poorly bonded inside, showing multiple vertical stripes. The annotation software called MAKE SENSE labels the images in the experiment according to the defect information. Label boxes are added, and the corresponding label files are generated for the areas with defects in images. A total of 223 original images have been acquired, and a total of 442 enhanced images are used in this experiment after the improved DCGAN is used to expand the image. The dataset images are then randomly divided into two groups: 80% of the dataset is used for parameter learning and network training, whereas the other 20% is used to test the generalization and recognition ability of the model, and the two datasets do not intersect each other.

B. EXPERIMENTAL ENVIRONMENT

The operating system of this experiment is Ubuntu 18.04.5 LTS, the GPU is GeForce RTX 2080 Ti, and the CPU is Intel(R) Xeon(R) CPU E5-2690 V3@ 2.60ghz. This experiment is improved on the basis of the YOLOv5s model. The framework is Pytorch, the number of training threads is 8, the batch size is 16, and the number of training epochs is 450. The sizes of nine groups of anchor boxes obtained by k-means algorithm clustering in this study are (15,30), (13,48), (19,38), (17,59), (14,81), (17,75), (24,55), (19,100), (14,144). The distribution is shown in Table 2.

TABLE 1. The table of defects number.

		Images	Labels	Debonding	Cracking	Ddelamination
Original images	The training set	179	287	246	16	25
	The test set	44	71	61	4	6
	Total	223	358	307	20	31
Expanded image	The training set	349	626	352	109	165
	The test set	93	157	90	28	39
	Total	442	783	442	137	204

TABLE 2. The distribution of anchor boxes.

Feature Map	8×8	16×16	32×32
Receptive Field	Large	Medium	Small
Anchor Boxes	(24,55),(19,100),(14,144)	(17,59),(14,81),(17,75)	(15,30),(13,48),(19,38)

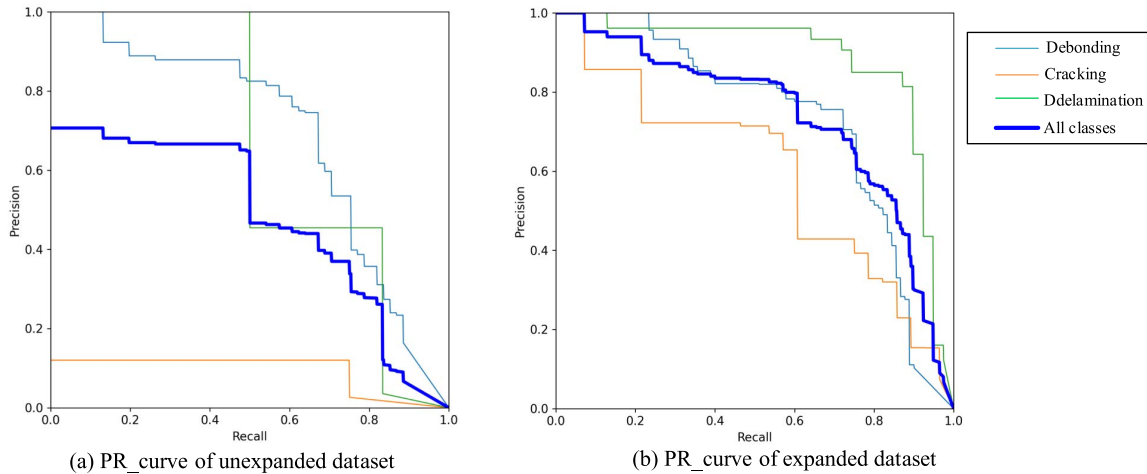


FIGURE 12. The trend of precision-recall curves.

C. EXPERIMENTAL ENVIRONMENT

In order to measure the robustness and accuracy of defect recognition, Precision, Recall, Average Precision (AP) at an IoU threshold of 0.5, and Mean Average Precision (mAP) are employed as the main evaluation indicators in this experiment.

Precision: measures the check accuracy of the model, that is, the probability of YOLOv5s predicts a certain category, and the precision is the proportion of correct detections in all prediction boxes.

Recall: measures the check-all rate of the model, that is, the probability that YOLOv5s is correctly classified into a certain category.

Precision and Recall are defined as follows, respectively:

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

where TP is the number of IoU > 0.5 between the predicted and truth boxes, FP is the number of IoU < 0.5 between the predicted and truth boxes, and FN is the number of missed real boxes.

AP: the area enclosed by the Precision-Recall curve and coordinate axes, which represents the effectiveness of the YOLOv5s network in detecting a certain category under different thresholds. Generally speaking, the higher the AP value

indicates the better target detection. The AP can be calculated by equation (15), where P(r) denotes the Precision-Recall curve.

$$AP = \int_0^1 P(r)dr \tag{15}$$

mAP: the average of AP for different categories, which is used to measure the detection effect of YOLOv5s network on all defect categories.

$$mAP = \frac{1}{|C|} \tag{16}$$

D. RESULTS AND DISCUSSION

1) COMPARISON EXPERIMENT OF DATA ENHANCEMENT

The image recognition effect of improved DCGAN for data enhancement is verified on the YOLOv5s model. Fig.12 shows the trend of Precision-Recall curves before and after data enhancement. Without any data enhancement method, the AP values of debonding, cracking, and delamination is 69.1%, 9.28%, and 65.2%, respectively. After using the improved DCGAN data enhancement method, the AP values are increased to 73%, 59.3%, and 87.9%, respectively, and the mAP value is increased by 25.5%.

As can be seen from Fig.13, the precision and recall of the YOLOv5 network are greatly increased after the dataset expansion, which fully indicates that the expansion of the dataset has a very important impact on the prediction of the

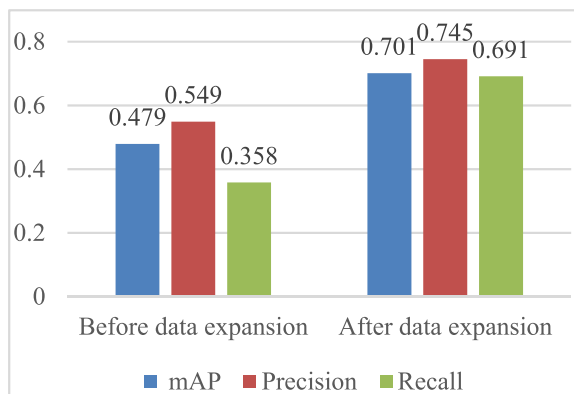


FIGURE 13. Comparison results before and after data expansion.

network. The YOLOv5s model has a significant improvement effect when using data enhancement method compared with no data enhancement method, indicating that data enhancement can effectively solve the problem of small and unbalanced number of images in the dataset and improve the generalization ability and robustness of the classification model.

2) COMPARISON EXPERIMENT OF LOSS FUNCTIONS

Analysis of Table 3, YOLOv5s has the best precision performance on GIOU, and has the best recall and mAP performance on EIOU. For mAP of different models, EIOU performs the best, followed by CIOU and DIOU, and GIOU has the worst performance. Although DIOU takes into account the shortcomings of GIOU, the aspect ratio of the bounding box is not considered in the regression process, and its precision is lower than that of GIOU. Since CIOU increases the loss of the detection box, the regression precision is improved on the basis of DIOU, and the prediction box is closer to the truth box. EIOU divides the loss term into the difference between the predicted width and height and the minimum width and height of the outer box on the basis of CIOU, which accelerates convergence and improves regression accuracy, and mAP and the recall rate reached maximum values. So our next aim is to improve the precision of YOLOv5s + EIOU through the CBAM attention mechanism.

3) EXPERIMENT ON CBAM ATTENTION MECHANISM

The four network models are compared and analyzed, and the corresponding training curve is drawn by tensorboard, as shown in Fig.14.

In the Fig.14, C3_1_CBAM and C3_2_CBAM indicate that CBAM is added after the C3 module of the backbone module and the neck module respectively, and C3_CBAM indicates that CBAM is added after all the C3 modules in the YOLOv5s network. During the network training process, the Box_Loss indicates whether an algorithm can locate the center point of an object well and whether the detection target is covered by the predicted bounding box. The smaller the loss

function value, the more accurate the prediction frame. The Cls_Loss represents the ability of the algorithm to correctly predict a given object category. The smaller the loss value, the more accurate the classification. The Obj_Loss is essentially a measure of the probability that the detection target exists in the region of interest. The smaller the value of the loss function, the higher the accuracy.

As shown in Fig.14, the loss function value has a downward trend during the training process, the Stochastic Gradient Descent algorithm optimizes the network, and the network weight and other parameters are constantly updated. Before the training epoch reached 300, the loss function value drops rapidly, and the precision, recall rate, and mAP rapidly improve. When the training epoch reaches approximately 300, the decrease in the loss function value gradually slowed. Similarly, the increases in the precision, recall rate and mAP also slowed. When the training epoch reached 430, the loss curves of the training showed almost no downward trends, and other index values also tend to have stabilized. The network model basically reached the convergence state, and the optimal network weight was obtained at the end of training. Fig.14(a) shows that after about 300 epochs of the YOLOv5s + EIOU + C3_CBAM model, the mAP reaches about 90%, and has gradually stabilized, reaching a maximum of 93.85%, indicating that the improved YOLOv5 model has an average precision rate for defect detection. Fig.14(b) shows that the precision reaches 83.09% when the YOLOv5s + EIOU + C3_CBAM model is trained to 300 rounds and continues to grow up to 87.93%. Fig.14(c) shows that the recall of the YOLOv5s + EIOU + C3_CBAM model is the first to slowly decline and then continue to grow to the highest value of 90.11%. The overall model performance has met and even exceeded expectations. The loss function can intuitively reflect whether the network model can converge stably as the number of iterations increases. The specific loss function of the model is shown in Fig.14(d), (e), (f) below. From the figure, it is found that as the number of epochs gradually increases, the YOLOv5s + EIOU + C3_CBAM algorithm curve gradually converges, and the loss value becomes smaller and smaller. When the model is iterated 430 times, the loss value is basically stable and has dropped to near 0, and the network basically converges. Compared with the other models, the YOLOv5s + EIOU + C3_CBAM model has better detection performance and recognition effect for adhesive structure defects, and the regression is faster and more accurate, which proves the validity and effectiveness of the model.

The above training model is used for testing, and the results are shown in Table 4.

The CBAM attention mechanism aims to improve the network's ability to extract important features, which is reflected in the result of an improvement of precision. As can be seen from the table, YOLOv5s + EIOU + C3_CBAM has the highest precision, followed by YOLOv5s + EIOU + C3_1_CBAM and YOLOv5s + EIOU + C3_2_CBAM, and YOLOv5s + EIOU is the lowest. Table 3 reveals the mAP

TABLE 3. Experimental results of different loss functions.

models	AP/%			mAP /%	Precision /%	Recall /%
	Debonding	Cracking	Ddelamination			
YOLOv5s + GIOU	70.9	54.5	84.9	70.1	74.5	69.1
YOLOv5s + DIOU	76.6	51.6	87.5	71.9	70.9	70.6
YOLOv5s + CIOU	73	59.3	87.9	73.4	72.7	74.2
YOLOv5s + EIOU	72.2	62.5	87.2	74	66.6	79.6

TABLE 4. Comparison table of experimental test results.

Models	mAP /%	Precision /%	Recall /%	Model size (MB)	Train Time (h)	Infer Time (ms)
1 YOLOv5s + EIOU	74	66.6	79.6	14.3	0.569	1.7
2 YOLOv5s + EIOU + C3_1_CBAM	78.4	75.8	73.6	14.4	0.598	2.6
3 YOLOv5s + EIOU + C3_2_CBAM	72.2	71.6	69.9	14.5	0.613	2.7
4 YOLOv5s + EIOU + C3_CBAM	78.6	77.2	76	14.6	0.670	2.9

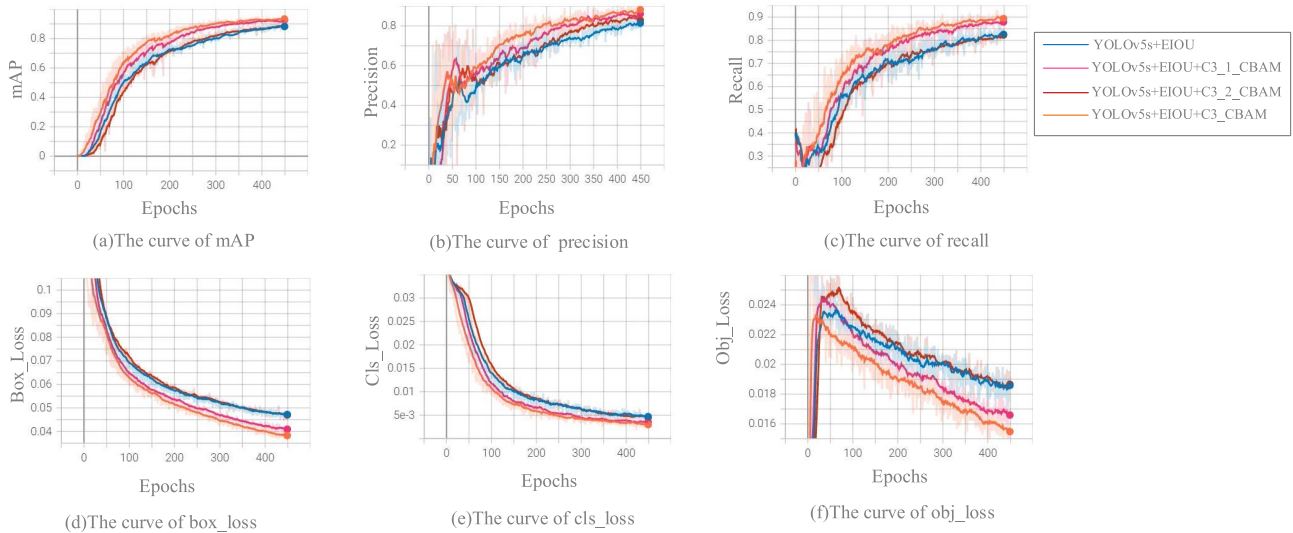


FIGURE 14. Training curves of different model.

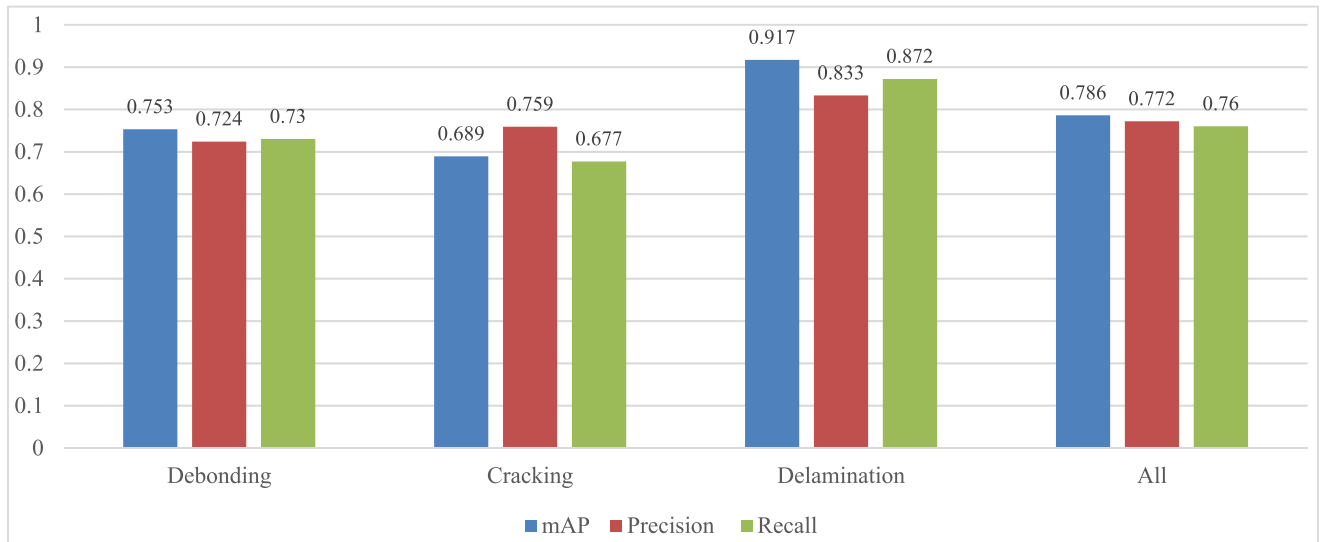
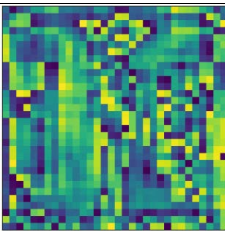
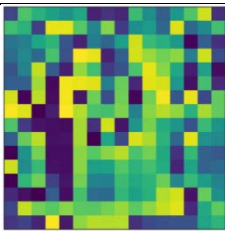
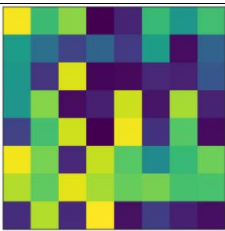
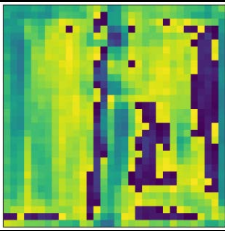
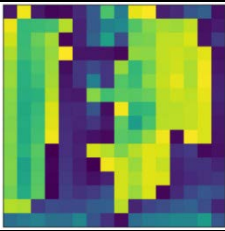
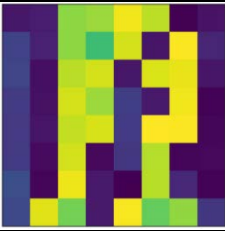
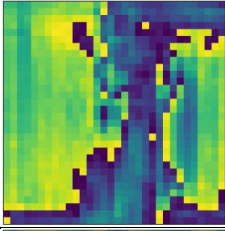
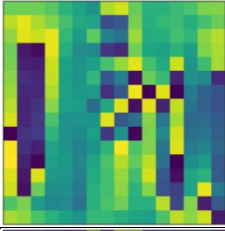
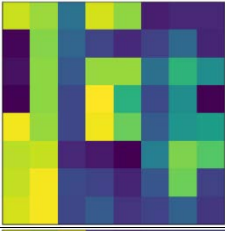
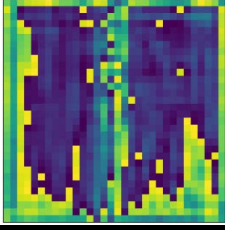
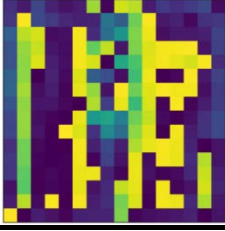
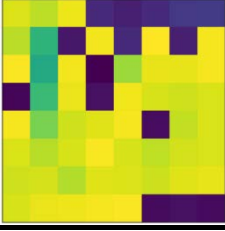


FIGURE 15. AP values for various defects.

and recall of YOLOv5s + EIOU were relatively good, but the precision had a bad advantage in the comparison experiment. Although YOLOv5s + EIOU + C3_1_CBAM's training time cost was higher than that of the YOLOv5s + EIOU

model, its mAP and precision were greatly improved compared with the original model. The mAP, precision, and recall rate of YOLOv5s + EIOU + C3_2_CBAM are lower than YOLOv5s + EIOU + C3_1_CBAM. Compared with

TABLE 5. Feature maps of the network model.

Models	32×32	16×16	8×8
YOLOv5s + EIOU			
YOLOv5s + EIOU + C3_2_CBAM			
YOLOv5s + EIOU + C3_1_CBAM			
YOLOv5s + EIOU + C3_CBAM			

YOLOv5s + EIOU, the precision of YOLOv5s + EIOU + C3_CBAM was increased by 10.6% and map by 4.6%, and the recall rate decreased slightly. Meanwhile, its detection precision, map, and recall rate were relatively balanced, and the detailed evaluation metrics are shown in Fig.15. However, the addition of CBAM mechanisms increased the depth of the network models, which sacrificed a certain speed advantage to improve the effect of detecting defects. The experimental results show that the in terms of the precision, recall rate, and map, the network improved based on YOLOv5s greatly outperformed the original YOLOv5s model, they are improved to 77.2%, 76%, and 78.6% respectively, thus capable of extracting features of defect adhesive structure images more accurately.

In order to better observe the output of each detection layer and the feature extraction effect of each layer of the model, the output layer feature maps of the above four models were visualized, as shown in Table 5.

The detection layer outputs feature maps of three scales, 32 × 32, 16 × 16, and 8 × 8, which are used to detect small, medium, and large objects. Among them, the 32 × 32 grid has a higher resolution and contains more location information, which is conducive to the location of defects,

and the 8 × 8 grid is obtained through a deeper network and contains more semantic information, which is conducive to defect classification. From the small target feature map (32×32), it can be seen that YOLOv5s + EIOU + C3_CBAM has a better visualization effect than other models, has a spatial correspondence with the original image, and has richer contour information. At this time, the receptive field is the smallest, which can improve the detection effect on small targets.

V. CONCLUSION

In this paper, the adhesive structure defect recognition method based on DCGAN and YOLOv5 is proposed. The DCGAN network is designed to expand the defect images and make the defect dataset, which solves the problem of few sample images and unbalanced distribution. According to the characteristics of adhesive structure defect image and YOLOv5 network, an improved YOLOv5s algorithm is proposed. It mainly focuses on the overall architecture design and optimization of the network, as well as the improvement of the loss function. The optimized YOLOv5s algorithm is then used to recognize the defect images. The comparison experiments show that the optimized YOLOv5s algorithm has

a better recognition effect than the pre-optimized YOLOv5s algorithm.

In practical industrial production, the use of deep learning often faces the problem of a small amount of raw data and the unbalanced distribution of defect samples in the dataset. To address this problem, we use the method of extending the dataset and optimizing the network model to make deep learning effective in practical applications. On this basis, it is worthwhile to further improve the precision and recall of the model, which is the direction of in-depth research.

REFERENCES

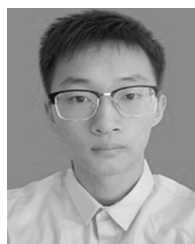
- [1] Z. Gu, X. Liu, and L. Wei, "A detection and identification method based on machine vision for bearing surface defects," in *Proc. Int. Conf. Comput., Control Robot. (ICCCR)*, Shanghai, China, Jan. 2021, pp. 128–132.
- [2] A. Tan, G. Zhou, and M. He, "Surface defect identification of citrus based on KF-2D-Renyi and ABC-SVM," *Multimedia Tools Appl.*, vol. 80, no. 6, pp. 9109–9136, Mar. 2021.
- [3] H. Xiao, D. Chen, J. Xu, and S. Guo, "Defects identification using the improved ultrasonic measurement model and support vector machines," *NDT E Int.*, vol. 111, Apr. 2020, Art. no. 102223.
- [4] C. Xu, L. Li, J. Li, and C. Wen, "Surface defects detection and identification of lithium battery pole piece based on multi-feature fusion and PSO-SVM," *IEEE Access*, vol. 9, pp. 85232–85239, 2021.
- [5] D. Pau, F. Previdi, and E. Rota, "Tiny defects identification of mechanical components in die-cast aluminum using artificial neural networks for micro-controllers," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2021, pp. 1–4.
- [6] J. Qi, X. Liu, K. Liu, F. Xu, H. Guo, X. Tian, M. Li, Z. Bao, and Y. Li, "An improved YOLOv5 model based on visual attention mechanism: Application to recognition of tomato virus disease," *Comput. Electron. Agricult.*, vol. 194, Mar. 2022, Art. no. 106780.
- [7] Z. Wang, L. Wu, T. Li, and P. Shi, "A smoke detection model based on improved YOLOv5," *Mathematics*, vol. 10, no. 7, p. 1190, Apr. 2022.
- [8] J. Yao, J. Qi, J. Zhang, H. Shao, J. Yang, and X. Li, "A real-time detection algorithm for kiwifruit defects based on YOLOv5," *Electronics*, vol. 10, no. 14, p. 1711, Jul. 2021.
- [9] D. Wang, E. Qin, and H. Yuan, "Classification of aquatic animals based on DCGAN-based data enhancement," *Fishery Modernization*, vol. 46, no. 6, pp. 68–75, Jul. 2019.
- [10] F. Gao, Y. Yang, J. Wang, J. Sun, E. Yang, and H. Zhou, "A deep convolutional generative adversarial networks (DCGANs)-based semi-supervised method for object recognition in synthetic aperture radar (SAR) images," *Remote Sens.*, vol. 10, no. 6, p. 846, May 2018.
- [11] L. Yifan, M. Yongzhi, and L. Banghuan, "Research on enhancement method of track defect sample based on deep convolution generative adversarial network," in *Proc. IEEE 2nd Int. Conf. Civil Aviation Saf. Inf. Technol. (ICCSIT)*, Wuhan, China, Oct. 2020, pp. 331–335.
- [12] L. Zhang, L. Duan, X. Hong, X. Liu, and X. Zhang, "Imbalanced data enhancement method based on improved DCGAN and its application," *J. Intell. Fuzzy Syst.*, vol. 41, no. 2, pp. 3485–3498, Sep. 2021.
- [13] Y. Cheng, Z. Dai, Y. Ji, S. Li, Z. Jia, K. Hirota, and Y. Dai, "Student action recognition based on deep convolutional generative adversarial network," in *Proc. Chin. Control Decis. Conf. (CCDC)*, Hefei, China, Aug. 2020, pp. 128–133.
- [14] W. Wang, W. Liu, J. Li, and W. Peng, "A rub fault recognition method based on generative adversarial nets," *J. Mech. Sci. Technol.*, vol. 34, no. 4, pp. 1389–1397, Apr. 2020.
- [15] B. Shi, X. Zhou, Z. Qin, L. Sun, and Y. Xu, "Corn ear quality recognition based on DCGAN data enhancement and transfer learning," in *Proc. 4th Int. Conf. Electron., Commun. Control Eng.*, Apr. 2021, pp. 62–68.
- [16] L. Zhao and R. Zhao, "Research on image inpainting based on generative adversarial network," in *Proc. Int. Conf. Comput. Netw., Electron. Autom. (ICNEA)*, Sep. 2020, pp. 259–263.
- [17] S. Liu, X. Wang, L. Wang, X. Zhang, and Z. He, "Abnormal behavior analysis strategy of bus drivers based on deep learning," in *Proc. IEEE 10th Data Driven Control Learn. Syst. Conf. (DDCLS)*, May 2021, pp. 1522–1527.
- [18] M. Li, H. Zhu, H. Chen, L. Xue, and T. Gao, "Research on object detection algorithm based on deep learning," *J. Phys., Conf.*, vol. 1995, no. 1, 2021, Art. no. 012046.
- [19] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Munich, Germany, 2018, pp. 3–19.



YONG JIN received the Ph.D. degree from the North University of China, in 2013. He is currently a Professor with the School of Information and Communication Engineering, North University of China. His research interests include image processing, online inspections, and big data analytics.



HUIFANG GAO is currently pursuing the master's degree with the School of Information and Communication Engineering, North University of China. Her research interests include deep learning and image processing.



XIAOLIANG FAN is currently pursuing the Ph.D. degree with the School of Earth Science and Engineering, Nanjing University. His research interests include soil desiccation cracking, machine identification, and self-healing of cracks.



HASSAN KHAN is currently pursuing the master's degree with the School of Information and Communication Engineering, North University of China. His research interests include image processing and online inspections.



YOUXING CHEN received the Ph.D. degree from the North University of China, in 2010. He is currently a Professor with the School of Information and Communication Engineering, North University of China. His research interests include image processing, signal processing, and non-destructive testing.

...