IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received 9 January 2024, accepted 4 March 2024, date of publication 7 March 2024, date of current version 14 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3374869

RESEARCH ARTICLE

Steel Surface Defect Detection Method Based on Improved YOLOX

CHENGFEI LI, AO XU, QIBO ZHANG, AND YUFEI CAI
Faculty of Intelligent Manufacturing, Wuyi University, Jiangmen, Guangdong 529020, China

Corresponding authors: Ao Xu (970305462@qq.com) and Chengfei Li (car1234566@sina.com)

This work was supported in part by the 2021 Graduate Education Innovation Plan Project of Guangdong Province of China under Grant 2021JGXM109, and in part by the 2022 Guangdong Undergraduate Colleges and Universities Teaching Quality and Teaching Reform Project Construction under Project GDJX2022005.

ABSTRACT Steel is a crucial material that is extensively utilized in various aspects of daily life and holds significant importance. However, during its production, there is a possibility of certain defects arising that could have a negative impact on the quality of the steel. Only by accurately detecting the defects on the steel surface can we avoid the harm caused by defects in steel. Because the steel surface defect detection algorithm is prone to misdetection, missed detection, and other problems, a steel surface defect detection algorithm based on improved YOLOX is proposed. First, the CSPCrossLayer module proposed in this paper is used to replace the CSPLayer structure in the backbone network to enrich the gradient information of the network and strengthen the feature extraction capability; Then, the SA (Shuffle Attention) module is added after the output of the backbone network, highlighting the general information to input high-quality features for the feature fusion network; Finally, the PSblock module is proposed to replace the CSPLayer structure in the feature fusion network, which reduces redundant computations to efficiently perform feature fusion on feature layers of different scales and improves the feature fusion capability of the model. The experiments involved testing the algorithm on two datasets: the publicly available NEU-DET dataset and a steel rail dataset collected in this paper. The algorithm can reach 77% mAP on the NEU-DET dataset, while the detection speed is 100 FPS. It reaches 88.8% mAP on the steel rail dataset, and the detection speed is 93FPS. These results demonstrate that the proposed algorithm is capable of swiftly and accurately detecting surface defects in steel materials.

INDEX TERMS Defect detection, YOLOX, gradient, shuffle attention.

I. INTRODUCTION

As one of the basic raw materials, steel is usually used in important industries, such as construction materials, machinery and equipment, and automobiles, so its quality must be guaranteed. In the process of continuous development of society, fields like aerospace and electronic industries have a high degree of refinement of steel. This makes steel companies control the quality of steel more and more strictly. The production process of steel mainly includes smelting, continuous casting, rolling, and quenching. Among them, smelting and continuous casting is the process of transforming raw iron into molten steel and then casting it into billet; rolling is the process of turning the billet into semi-finished products of different specifications and shapes through hot rolling, cold rolling, etc.; and finally quenching to get the specific performance of the steel. But in rolling and some other processes, some technical problems and environmental factors will inevitably lead to many kinds of defects, such as crazing, inclusion, and rolled-in scale. These defects will not only affect the quality and life of steel but also cause harm in actual use. Therefore, rapid and accurate detection of steel surface defects is a very important part of the process.

In industrial production, manual inspection methods are widely used for the detection of steel surface defects [1]. These manual inspection methods rely on visual observation and experience to accurately detect steel surface defects. However, the method is inefficient, labor-intensive, and

The associate editor coordinating the review of this manuscript and approving it for publication was Wei-Yen Hsu.

expensive, and artificially long hours of labor will lead to misdetection and missed detection of the situation. Defect detection algorithms are evolving as a result of the constant advancement of science and technology. These algorithms are primarily separated into defect detection algorithms based on traditional machine learning and defect detection algorithms based on deep learning. Defect detection algorithms based on traditional machine learning are essentially classified into three types [2]: methods based on texture features [3], [4], methods based on color features [5], and methods based on shape features [6]. These algorithms are mainly composed of feature extraction (HOG [7], SIFT [8], etc.) and classifiers (SVM [9], etc.). These methods generally need to manually select the features used to distinguish between defective areas and other areas, relying heavily on the expert's knowledge and judgment. Most of the extracted features are shallow features, which are difficult to detect in complex scenarios. These methods also have poor generalization and are difficult to use in actual production. Unlike traditional machine learning algorithms that require manual design of feature extraction rules, deep learning-based detection algorithms automatically extract deeper features through convolution and other operations, with higher accuracy, applicability, and robustness.

With the fast advancement of deep learning in recent years, target detection algorithms based on deep learning have been endless, mainly divided into one-stage networks represented by SSD [10], CenterNet [11], and YOLO series [12], [13], [14], [15] and two-stage networks represented by R-CNN [16], Fast R-CNN [17], Faster R-CNN [18], and Mask R-CNN [19]. The main difference between the two is whether or not there is a phase for generating regional candidate boxes. The one-stage target detection algorithm does not need to generate the region candidate box. Detection results can be directly calculated through the network, which is fast, but the detection accuracy may be relatively low. The two-stage target detection algorithm process is divided into two phases. Firstly, generating candidate boxes, and then refining the detection points based on these candidates for higher accuracy, but at the cost of slower detection speed. The detection accuracy is higher, but the detection speed is slow. Among them, the YOLO series of algorithms in terms of detection speed and accuracy is very balanced and better suited for detecting industrial defects [20].

YOLOX [21] is a one-stage target detection algorithm proposed by Megvii in 2021. Unlike the previous YOLO series algorithms that use an anchor-based approach, YOLOX introduces an anchor-free mechanism to reduce the number of generated prediction frames; it proposes a fast-converging decoupled detection header and SimOTA dynamic positive sample allocation. At the same time, Mosaic and Mixup data enhancement methods are performed on the input image to increase the diversity of the data set and improve the generalization ability of the model. Through these upgrades, YOLOX's detection accuracy and inference speed have both increased, making it better suited for the detection of industrial defects. In this paper, based on YOLOX-s, we improve it for the problems of low accuracy of steel surface defect detection and easy to miss and misdetect, and we mainly have the following contributions:

1. To improve the quality of the feature layers entering the feature fusion network, the SA (Shuffle Attention) module is added between the backbone network and the feature fusion network to highlight important information while suppressing unnecessary information;

2. To address the issue of the complicated texture of steel surface defects, which is challenging to extract and result in missed detection, the CSPCrossLayer module is used to enrich the gradient in the backbone network to improve the parameter utilization rate of the model and strengthen the feature extraction capability of the backbone network;

3. To efficiently acquire feature information and enhance the feature fusion capability, the PSblock module is used in the feature fusion network to reduce the redundant computation to adequately fuse feature layers of different scales.

The remainder of this paper is composed as follows. Section II introduces some related work on surface defect detection of steel materials; Section III mainly describes the method used in this paper; Section IV mainly introduces the public dataset and the self-collected rail dataset used in this paper, and analyzes the effectiveness of the algorithm through ablation experiments and comparative experiments. Finally, the conclusion is presented in Section V.

II. RELATED WORK

With the swift advancement of deep learning, a proliferating mass of people use deep learning target detection algorithms for defect detection. Defect detection algorithms based on deep learning can be categorized into one-stage algorithms and two-stage algorithms. The two-stage algorithm represented by Faster-RCNN first generates candidate frames and then performs classification and regression. Wei et al. [22] proposed several strategies to further enhance Faster-RCNN to address issues such as the misdetection of small objects and the diversity of defect types. These improvements include weighted ROI pooling, a multi-scale feature extraction network using FPN, and a strict-non-maximum suppression (Strict-NMS) algorithm, which aim to enhance the detection performance of the model. Zhao et al. [23] presented an improved Faster-RCNN network model, which first replaces part of the regular convolution with variability convolution to reconstruct the ResNet50 for feature extraction. Secondly, they utilized a feature pyramid network to merge multiscale features and replaced fixed ROI pooling with deformable pooling. Lastly, the soft non-maximal suppression algorithm (SNMS) is employed to suppress detection boxes with significant overlap with the highest-scoring box, thereby enhancing the network's ability to recognize surface defects in steel materials. Wang et al. [24] proposed an enhanced Faster-RCNN network that combines improvements with the ResNet50 architecture. They introduced a deformable rotating network within the ResNet50 to enhance the detection

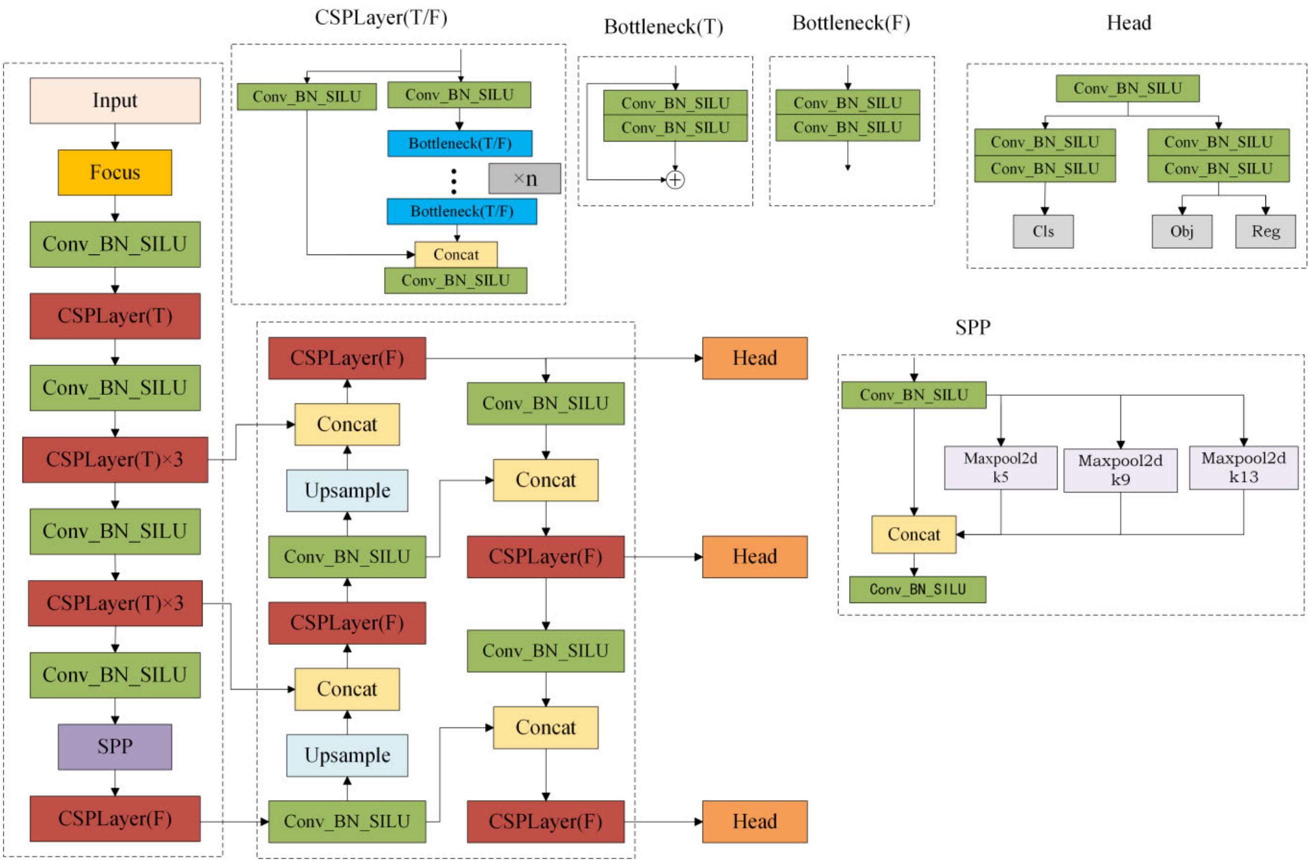


FIGURE 1. YOLOX structure.

capability for defects of different shapes. Additionally, they incorporated spatial pyramid pooling (SPP), enhanced feature pyramid (FPN), and matrix NMS algorithms to improve the detection accuracy of the model.

The two-stage network has a higher detection accuracy but a slower detection speed. The one-stage network has a simple structure, and the results are calculated directly through the network, which can ensure the balance of detection accuracy and speed at the same time. Classic one-stage detection networks include SSD, YOLO series, CenterNet, etc. Tian et al. [25] proposed a steel surface defect detector called DCC-CenterNet. They primarily introduced a dilated feature enhancement model (DFEM), a new center-weight function, and a CIOU loss function to improve detection accuracy. The detection speed of DCC-CenterNet exceeds 70 FPS, meeting the requirements for real-time detection. Li et al. [26] made improvements to the YOLOv4 model. They designed a convolutional attention module for the backbone network and replaced the augmented path aggregation network (PANet) with a receptive field block (RFB) to enhance the network's feature extraction capability and achieve higher detection accuracy. Wang et al. [27] developed a steel surface defect detection technique based on the YOLOv5. They introduced a multi-scale module and spatial attention mechanism to enhance detection performance, ensuring real-time speed requirements. Ge et al. proposed YOLOX, a one-stage end-to-end object detection algorithm that achieves a better balance between detection accuracy and speed. Building upon this, this paper presents a steel surface defect detection method based on YOLOX.

III. METHOD

A. YOLOX OVERVIEW

1) NETWORK STRUCTURE

As shown in Fig. 1, the YOLOX structure is mainly divided into three parts: the backbone network, the feature fusion network, and the detection head. The input image goes through the backbone network to obtain the feature maps with down-sampling multiples of 8, 16, and 32, and then these feature maps go through the feature fusion network for feature fusion. Finally, the coordinates, border size, category, and confidence level of the object are predicted by the detection head part.

The backbone network uses CSPDarknet as the feature extraction network. It comprises several key structures, such as the focus structure, residual structure, CSPLayer structure, and SPP structure, which collectively contribute to effective extraction.

The use of the residual structure with jump connections mitigates the problem of vanishing gradients brought about by increasing the depth of the network.

CSPLayer network structure: the residual fast is split into two parts: one part carries out the stacking of the original residual fast, and the other part serves as a residual edge, which is finally spliced.

Focus structure: Four independent feature layers are obtained by taking out a value at every other pixel point on the feature map, and then these feature layers are stacked in the channel dimension. The width and height of the final feature layer are reduced to one-half of the original, and the number of channels is expanded to four times the original. This is shown in Fig. 2.

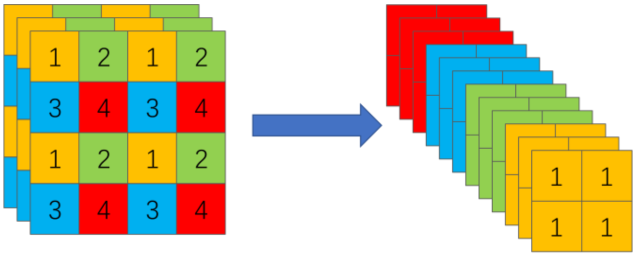


FIGURE 2. Focus structure.

SPP structure: spatial pyramid pooling structure, which is pooled by pooling kernels for the largest pooling layers of 5, 9, and 13, and finally spliced to increase the sensory field of the network.

The feature fusion network uses a path aggregation network (PANet), a two-way feature fusion where first the high-level features are fused to the low-level features, after which the low-level features are then fused to the high-level features for better feature fusion.

In the detection head section, YOLOX introduces a novel decoupled detection head that separates the classification and regression tasks. This method speeds up the model convergence and increases the detection performance.

2) PRINCIPLE OF PREDICTION

Taking 256×256 input as an example, data augmentation techniques such as mosaic and mixup are applied before inputting into the neural network, and then feature extraction is carried out through the backbone network. The input image is subjected to five downsampling operations, and after the last downsampling, the sensory field of the model is enlarged through the SPP pooling module. The feature layer with downsampling multiples of 8, 16, and 32 is taken into the PAFPN structure for feature fusion, outputting feature maps with sizes of 32×32, 16×16, and 8×8. Finally integrated through the head output layer to achieve classification and regression on targets of different sizes.

3) LOSS FUNCTION

The loss function of YOLOX is divided into three components: prediction frame loss (L_{reg}), confidence loss (L_{conf}), and classification loss (L_{cls}). The bounding box loss is calculated using the IOU loss function, while the confidence loss and classification loss are computed using the binary cross-entropy loss function. The formula is as follows:

$$Loss = L_{reg} + L_{conf} + L_{cls}, \tag{1}$$

$$L_{reg} = -ln\frac{BOX_{\cap}}{BOX_{\cup}}, \tag{2}$$

$$L_{cls} = -\sum\nolimits_{i=0}^{n} \left(t_i log\left(p_i\right) + \left(1 - t_i\right) log\left(1 - p_i\right)\right). \tag{3}$$

where t_i is the true category label and p_i is the probability of network prediction.

B. SA MODULE

In steel surface defect images, apart from the necessary identification of defects, there is also complex background information. This background information can interfere with the model's learning ability, thus affecting its accuracy and performance in detecting defects. The introduction of the attention mechanism can effectively address this issue. The attention mechanism helps improve the expressive power of the model for feature representation, and its main role is to highlight important information while suppressing unnecessary information, thereby helping the object detection network achieve better localization and recognition [28]. Currently, attention mechanisms are commonly categorized into two groups: channel attention mechanisms [29], [30], [31], [32] and spatial attention mechanisms [33], [34], [35], from the channel and pixel levels, respectively, to strengthen the feature map. The combination of these two types of attention mechanisms can achieve better results while increasing the amount of computation.

The SA (Shuffle Attention) [36] module ensures less computation while combining channel attention and spatial attention, and the structure is shown in Fig. 3.

As shown in Fig. 3, for the input feature graph X, the SA module first divides it into g groups of sub-features in the channel dimension. Then, each sub-feature is further sliced into two parts, which are used for channel attention and spatial attention computation to generate the corresponding importance coefficients. Finally, all the sub-features are pooled together to allow the features of different groups to interact with each other using the channel shuffle operation.

Channel attention is considered for lightweight design using only global pooling, parameter scaling, and sigmoid operations, which are calculated as follows:

$$F_{gap} = \frac{1}{H \times W} \sum\nolimits_{i=1}^{H} \sum\nolimits_{j=1}^{W} X_{s1}(i,j), \tag{4}$$

$$X'_{s1} = \sigma(F_c(F_{gap}))X_{s1} = \sigma(w_1 \cdot F_{gap} + b_1)X_{s1}. \tag{5}$$

where X_{s1} denotes the feature map for performing channel attention computation and σ is the sigmoid function.

The group norm is used for spatial attention, with the same parameter scaling and final sigmoid operation.

$$\begin{aligned} X'_{s2} &= \sigma\left(F_c\left(GN\left(X_{s2}\right)\right)\right) \\ &= \sigma\left(w_2 \cdot GN(X_{s2}) + b_2\right) X_{s2}. \end{aligned} \tag{6}$$

This attention mechanism is incorporated after the three different scales of feature layers output by the backbone

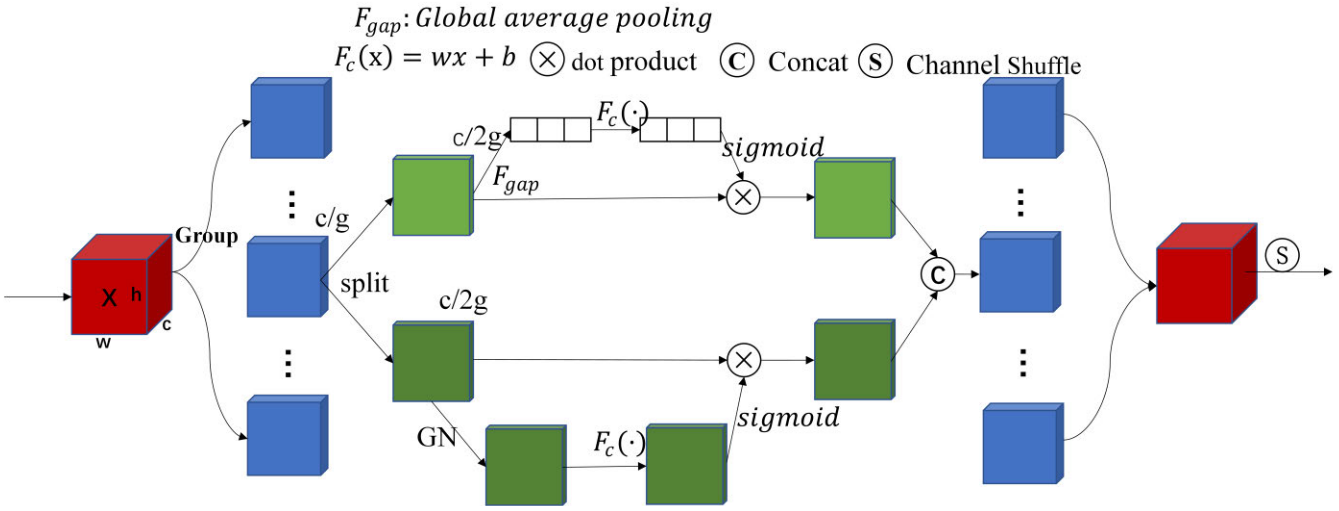


FIGURE 3. SA module.

network. This allows the network to reorganize the information using the attention mechanism before performing feature fusion. By highlighting important information and suppressing irrelevant details, the attention mechanism enhances the capability of the network to perform effective feature fusion.

C. CSPCROSSLAYER MODULE

Steel surface defects have the problem of similarity between defects and background, which requires the network to have stronger feature extraction capability. Since the learning of parameters in the network is realized by backpropagation, by adjusting the propagation path [37], different computing units can be made to learn richer information and increase the parameter utilization of the network to strengthen the learning ability of the network. At the same time, the rich gradient paths can avoid the degradation of the network's performance during training, so that the network has a stable learning ability.

In this paper, we propose the gradient-enriched CSPCrossLayer module, which introduces cross-layer connections to the cat splicing operation after the first convolutional layer in the backbone of the CSPLayer structure and after each bottleneck block. This module is used to replace the CSPLayer structure in the backbone network. The structure is shown in Fig. 4. Due to the addition of some gradient paths to the original structure, a small number of parameters and computations are added compared to the original structure. However, the utilization of the model parameters is improved, which gives the network a stronger feature extraction capability.

D. PSBLOCK MODULE

In the production of a variety of defects, some of the defects are very similar to each other and difficult to distinguish, and the detection process requires a model with stronger discriminatory ability. Better utilization of feature information to reduce unnecessary redundant computation can make the network have an efficient learning ability with stronger discriminative ability.

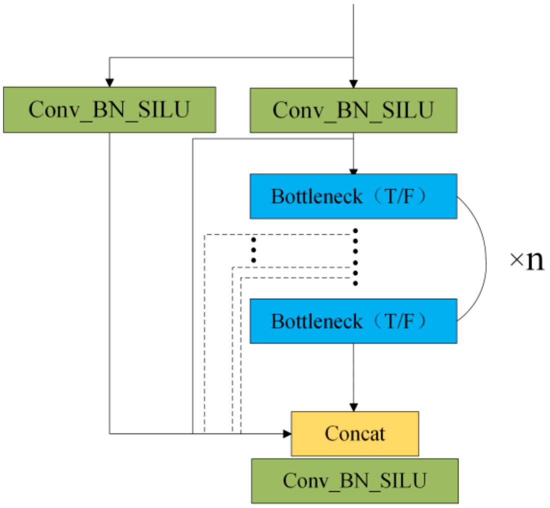


FIGURE 4. CSPCrossLayer structure.

Partial convolution (PConv) [38] is a lightweight convolution approach. This convolution approach can effectively reduce the redundant computation in the convolution operation to better utilize the feature layer. Its structure is shown in Fig. 5. Partial convolution (PConv) performs a convolution operation on only a part of the input channel and does not perform any operation on the other channels.

To reduce redundant computation and enhance the feature interaction capability of the feature fusion network, this paper proposes the PSblock module to replace the CSPLayer module in the feature fusion network, and the structure is shown in the following figure.

PSblock first consists of a residual structure consisting of a PConv and a 1×1 convolution, and finally a channel shuffle operation (Channel Shuffle). The PConv is used to reduce unnecessary redundant computations, since the PConv

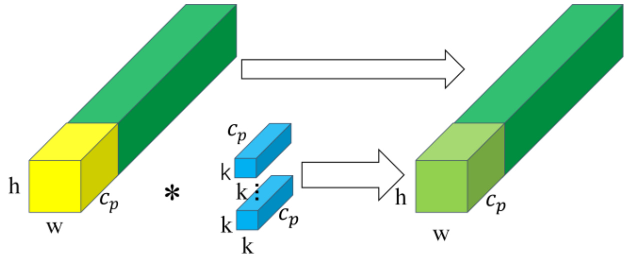


FIGURE 5. PConv structure.

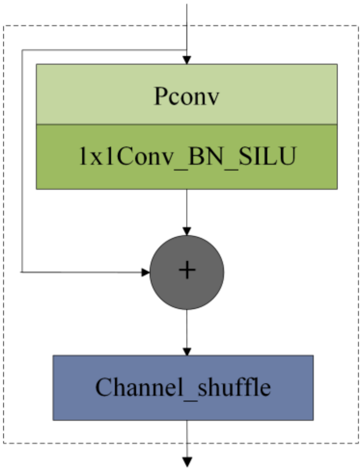


FIGURE 6. PSblock structure.

only convolves some channels and leaves the other channels unchanged, which leads to a lack of communication between these channels. Here, a 1×1 convolution is used to allow the channels to communicate with each other after the PConv and to be fused. Then the residual structure is used to further enrich the fused information of the channels to achieve a stronger feature fusion. Finally, the channel shuffle operation is used to further realize the interaction between channels by randomly reorganizing the output channels to improve the performance of the model.

IV. EXPERIMENT

A. DATA SET

To verify the effectiveness of the improved model, this study performed validation on two datasets. The first dataset is the NEU-DET [39] dataset, released by Northeastern University, which has 1800 pictures divided into 6 types of defects with 300 pictures in each type. The defect types include crazing, inclusion, patches, scratches, pitted surface, and rolled-in scale. Each image in the dataset has a size of 200×200 pixels. The specific pictures are shown below.

The other dataset is the steel rail dataset collected in this paper. The dataset was collected at the factory site. To avoid issues such as glare during the sampling process, we used tunnel lighting and industrial cameras to capture defect samples on the surface of the steel rails. The collection equipment is shown in Fig. 8. Blue light has a shorter wavelength, a weaker diffraction effect, a stronger ability to engrave details, and the ability to get a picture of a surface with more complete surface details, so we chose a blue light source to illuminate the metal.

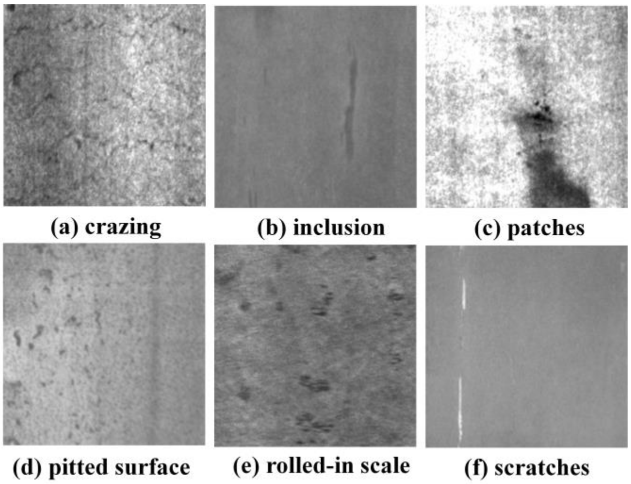


FIGURE 7. Dataset examples of NEU-DET.

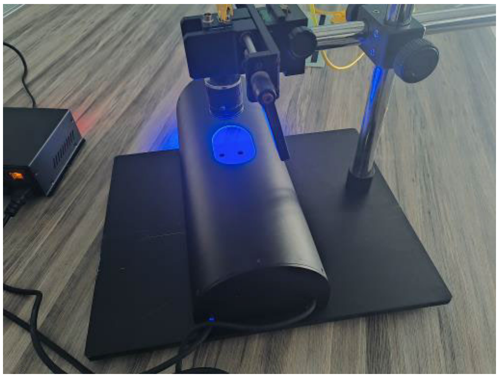


FIGURE 8. Collection equipment.

This dataset has a total of 2,120 images of steel rail defects, which are categorized into six categories: burn (Bu), dense rust (Dr), not polished (Np), pitting (Pi), scratches (Sc), burn (Bu), and steps (St). The specific number of each defect category is shown in Table 1. The image size of the dataset in this paper is 384×384, and the images of each category are shown in Fig. 9.

TABLE 1. Number of defects by category.

Categories	Bu	Dr	Np	Pi	Sc	St
Numbers	364	400	330	300	326	400

B. EVALUATION INDICATORS

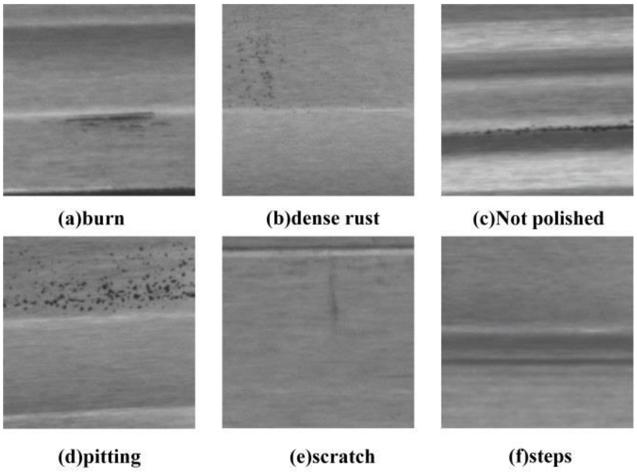
The following indicators are introduced to evaluate the model's performance: mean average precision (mAP), average precision (AP), and FPS (frames per second), which are formulated below.

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

$$R = \frac{TP}{TP + FN}, \tag{8}$$

TABLE 2. Ablation experiment.

Model	S	C	P	NEU-DET mAP(%)	NEU-DET FPS	Steel Rail mAP	Steel Rail FPS	Params	FLOPs(G)
Model 1				73.8	90	87.3	88	804251	21.6
Model 2	√			74.2	83	87.7	82	804287	21.6
Model 3		√		74.4	88	88.0	86	829851	22.4
Model 4			√	74.6	105	87.8	101	697527	19.8
Model 5	√	√		75.6	82	88.2	81	829887	22.4
Model 6	√		√	76.0	92	88.2	90	697563	19.8
Model 7		√	√	75.1	101	88.2	99	723127	20.7
Model 8	√	√	√	77.0	100	88.8	93	723163	20.7



(a)burn (b)dense rust (c)Not polished

(d)pitting (e)scratch (f)steps

FIGURE 9. Dataset examples of steel rail.

$$AP = \int_0^1 P(R)dR, \tag{9}$$

$$mAP = \frac{\sum_{i=0}^{n} AP(i)}{n}. \tag{10}$$

Where P and R denote precision and recall, respectively, TP represents the number of true positive predictions where the actual and predicted classes are both positive. FP represents the number of false positive predictions where the actual class is negative but predicted as positive. FN represents the number of false negative predictions where the actual class is positive but predicted as negative.

C. EXPERIMENTAL ENVIRONMENT

This experiment is completed under the Windows 10 system environment, the graphics card used is an NVIDIA GeForce RTX3080, and the CPU uses an Intel(R) Core(TM) i9-10850K@3.60 GHz. In both the NEU-DET dataset and the dataset collected in this study, the ratio of the training-validation set to the test set is 4:1. Within the training-validation set, the ratio between the training set and validation set is 4:1. To ensure the reliability of the experiments, the input size is uniformly adopted as 256×256 in the NEU-DET dataset and 384×384 in the dataset collected in this paper, and the number of iterations of the model is 350 epochs. The Adam optimizer is used, the initial learning rate is set to 0.001, the learning rate decay is adopted by the cosine annealing algorithm, and the batch size is set to 16.

D. EXPERIMENT ANALYSIS

1) ABLATION EXPERIMENT

To verify the effectiveness of the proposed modules in this paper, eight sets of ablation experiments are set up, and the experimental environment settings are kept the same. The results of the ablation experiments are shown in Table 2, where "$\surd$" indicates that the corresponding improvement method is adopted, S denotes the SA (Shuffle Attention) module, C denotes the CSPCrossLayer module, and P denotes the PSblock module, which are analyzed as follows:

a. The first group is the experimental results of the YOLOX-s algorithm before improvement. Using this as a baseline comparison, its detection mAP is 73.8% in the NEU-DET dataset and 87.3% in the steel rail dataset;

b. The second group is the experiment of adding the SA module, which improves the mAP by 0.4% in the NEU-DET dataset and 0.4% in the steel rail dataset without increasing the computation;

c. The third group is to replace the CSPLayer module in the backbone network with the CSPCrossLayer module with richer gradients. This module increases the computation slightly but improves the mAP by 0.6% in the NEU-DET dataset and 0.7% in the steel rail dataset;

d. The fourth group replaces the CSPLayer module in the feature fusion network with the PSblock module. By reducing redundant computation in the feature fusion process to fuse features more efficiently, which reduces the computation of the model and improves its performance. The mAP is improved by 0.8% in the NEU-DET dataset and 0.5% in the steel rail dataset;

e. The fifth, sixth, and seventh experiments use the proposed two improvement points simultaneously, and any two improvement points in the two datasets have different degrees of improvement. The eighth group uses three improvement points at the same time. Compared with the YOLOX-s algorithm without any improvement, it has a 3.2% mAP improvement in the NEU-DET dataset and a 1.5% mAP improvement in the steel rail dataset.

TABLE 3. Comparative experiments on the NEU-DET dataset.

Model	Crazing	Inclusion	Patches	Pitted Surface	Rolled-in Scale	Scratches	mAP	FPS
SSD	52.0	82.0	89.0	71.0	61.0	56.0	68.6	121
Faster-RCNN	35.3	69.1	80.6	80.6	37.9	77.5	63.5	60
CenterNet	48.0	78.0	90.0	83.0	58.0	75.0	71.9	102
YOLOv4	52.0	79.5	94.2	80.7	42.5	74.0	70.5	57
PPYOLOe-s [40]	58.0	78.1	92.5	81.5	47.3	77.8	72.5	72
YOLOv5-s	50.8	78.6	94.5	83.2	54.4	75.9	72.9	108
YOLOV7 [41]	53.8	80.7	93.6	81.6	46.8	84.3	73.5	96
YOLOX-m	55.5	81.2	93.0	81.4	52.7	86.3	75.0	66
OURS	55.1	83.0	93.6	86.1	59.7	84.2	77.0	100

TABLE 4. Comparative experiments on steel rail datasets.

Model	Burn	Dense Rust	Not Polished	Pitting	Scratches	Steps	mAP	FPS
SSD	86.0	87.0	94.0	59.0	58.2	93.0	79.5	123
Fater-RCNN	88.7	87.1	90.9	59.2	59.3	90.0	79.2	46
CenterNet	91.0	89.0	98.0	58.0	60.0	100	82.7	96
YOLOv4	76.8	82.6	93.2	58.4	46.9	98.2	76.0	55
PPYOLOe-s	83.7	81.2	97.4	64.8	41.6	98.7	77.9	67
YOLOv5-s	82.3	82.3	97.6	58.9	44.0	99.5	77.4	106
YOLOV7	85.8	88.4	94.2	67.1	60.4	96.6	82.1	93
YOLOX-m	95.2	91.9	99.0	68.5	73.5	99.5	87.9	65
OURS	93.5	89.5	99.2	72.1	79.2	99.5	88.8	93

2) COMPARISON EXPERIMENT

To further verify the effectiveness and superiority of the improved algorithm proposed in this paper, the algorithm of this paper is experimentally compared with other classical algorithms in the NEU-DET dataset under the guarantee of constant hyperparameters. The results are shown in Table 3. The results show that the improved YOLOX algorithm proposed in this paper has obvious advantages in the mAP indexes while ensuring lightweight. It performs the best on defects such as inclusions, pitted surfaces, and rolled-in scales, and the FPS reaches 100, which achieves a better balance between detection accuracy and inference speed.

The practicality of the algorithm still needs to be verified in conjunction with the actual field. The algorithm of this paper and the above classical algorithm are compared experimentally in the collected steel rail dataset. The experimental results are shown in Table 4.

As seen from the table, the improved YOLOX algorithm is still the best in terms of the performance of the mAP index, which is the best performance in the not polished, pitting, and scratches. The detection speed of the algorithm reaches 93 FPS, which meets the requirements of real-time detection in industrial production and has a certain practical value.

E. MODEL VALIDATION ANALYSIS

To visualize the effect of the improved algorithm, some pictures from the test set of the public dataset NEU-DET are extracted to use the improved algorithm and the YOLOX-s algorithm for detection and demonstration, respectively. At the same time, compared with the original labeled information. The effect of the algorithm before and after the improvement is shown in Fig. 10.

The comparison of the figures reveals that the accuracy of the enhanced algorithm has been significantly improved. In the original YOLOX algorithm, there were instances of missed detection in these sample images. However, our algorithm addresses this issue and successfully avoids such occurrences, resulting in a significant improvement in both steel surface defect misdetection and missed detection. Moreover, it enables more confident identification of defects, further confirming the effectiveness of the improved algorithm.

F. DISCUSSION OF EXPERIMENTAL RESULTS

This paper focuses on the improvement of the existing algorithms for challenges such as the many types of defects on the steel surface, the similarity between defects and backgrounds, and the ease of misdetection and missed detection. The effectiveness of the proposed improvement points is verified by ablation experiments. It shows good results in both the NEU-DET dataset and the steel rail dataset, which can detect defects on steel surfaces well. It effectively improves the problems of misdetection and missed detection in the detection of steel surface defects. It also effectively reduces the number of parameters in the model by about 10%, improves the detection speed of the model, and meets the real-time detection requirements.

Although the work in this paper has had some success, there is still much room for improvement:

a. In the actual production process, there are often some defects with much less sample data than others. For such defects, it is difficult to train for detection. To address this problem, small sample learning methods can be considered to detect defects of the same kind through a small number of samples.

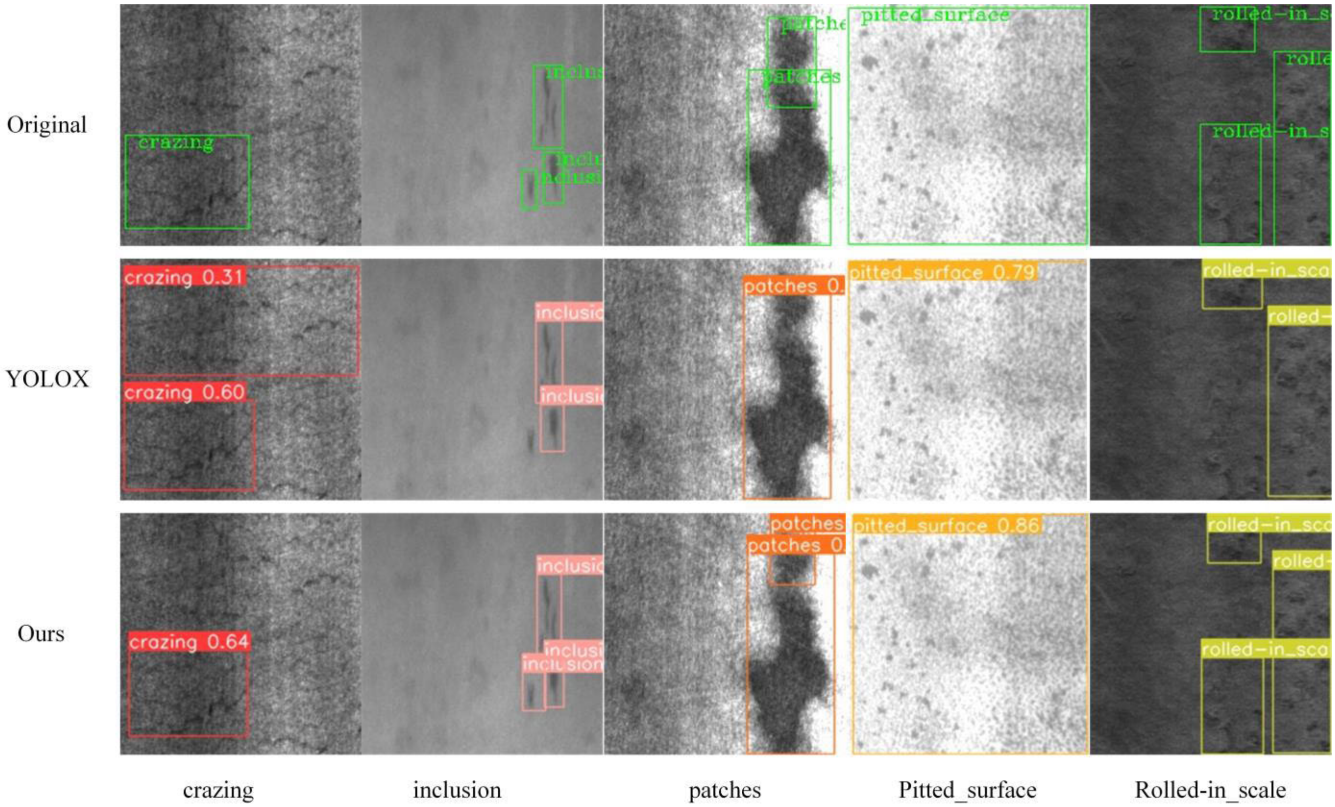


FIGURE 10. Comparison of algorithm effectiveness.

b. There are deficiencies in the detection of small targets, and in the future, we can consider reducing the information loss when fusing feature maps of different sizes to improve the effect of detecting small targets.

V. CONCLUSION

This paper proposes a steel surface defect detection method based on the improved YOLOX network for the problems of low detection accuracy, easy misdetection, and missed detection. The improved CSPCrossLayer module with richer gradients is used to replace the CSPLayer structure in the backbone network to improve the parameter utilization of the network and strengthen the feature extraction capability; the SA attention module is added behind the output of the backbone network to highlight the important features to improve the quality of the feature layer to enable better fusion; and the PSblock module is used to replace the CSPLayer structure in the feature fusion network to reduce redundant computation and improve the feature fusion capability.

The experimental results verify the effectiveness of each improvement point, and the final algorithm achieves 77% and 88.8% mAP in the NEU-DET dataset and the steel rail dataset. Compared with the original YOLOX algorithm, the improvement is 3.2% and 1.5%, respectively, and the inference speed of the improved algorithm in the two datasets is 100FPS and 93FPS, respectively, which meets the demand for industrial inspection.

The steel surface has numerous small defects, and a high downsampling factor under the YOLO algorithm easily results in the loss of information regarding these small target defects. The next step of our work focuses on optimizing the algorithm to address the issue of small targets while ensuring real-time detection requirements.

REFERENCES

[1] C. Li, H. Liang, S. Qiu, and A. Xu, "Steel surface defect detection method and application based on improved CenterNet," in *Proc. China Autom. Congr. (CAC)*, Xiamen, China, Nov. 2022, pp. 386–389.

[2] X. Wen, J. Shan, Y. He, and K. Song, "Steel surface defect recognition: A survey," *Coatings*, vol. 13, no. 1, p. 17, Dec. 2022, doi: 10.3390/coatings13010017.

[3] M. Chu, A. Wang, R. Gong, and M. Sha, "Strip steel surface defect recognition based on novel feature extraction and enhanced least squares twin support vector machine," *ISIJ Int.*, vol. 54, no. 7, pp. 1638–1645, 2014, doi: 10.2355/isijinternational.54.1638.

[4] Y.-J. Jeon, D.-C. Choi, J. P. Yun, and S. W. Kim, "Detection of periodic defects using dual-light switching lighting method on the surface of thick plates," *ISIJ Int.*, vol. 55, no. 9, pp. 1942–1949, 2015, doi: 10.2355/isijinternational.isijint-2015-053.

[5] Y. Wang, H. Xia, X. Yuan, L. Li, and B. Sun, "Distributed defect recognition on steel surfaces using an improved random forest algorithm with optimal multi-feature-set fusion," *Multimedia Tools Appl.*, vol. 77, no. 13, pp. 16741–16770, Jul. 2018, doi: 10.1007/s11042-017-5238-0.

[6] H. Hu, Y. Li, M. Liu, and W. Liang, "Classification of defects in steel strip surface based on multiclass support vector machine," *Multimedia Tools Appl.*, vol. 69, no. 1, pp. 199–216, Mar. 2014, doi: 10.1007/s11042-012-1248-0.

[7] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 886–893.

[8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004, doi: 10.1023/b:visi.0000029664.99615.94.

[9] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/bf00994018.

[10] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A.-C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands, 2016, pp. 21–37.

[11] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," 2019, *arXiv:1904.07850*.

[12] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[13] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525.

[14] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[15] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[16] R. Girshick, J. Donahue, T. Darrell, U. Berkeley, and J. Malik, "R-CNN: Region-based convolutional neural networks," in *Proc. Comput. Vis. Pattern Recognit.*, 2014, pp. 2–9.

[17] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[18] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.

[19] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[20] C. Li, J. Cai, S. Qiu, and H. Liang, "Defects detection of steel based on YOLOv4," in *Proc. China Autom. Congr. (CAC)*, Beijing, China, Oct. 2021, pp. 5836–5839.

[21] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021, *arXiv:2107.08430*.

[22] R. Wei, Y. Song, and Y. Zhang, "Enhanced faster region convolutional neural networks for steel surface defect detection," *ISIJ Int.*, vol. 60, no. 3, pp. 539–545, 2020, doi: 10.2355/isijinternational.isijint-2019-335.

[23] W. Zhao, F. Chen, H. Huang, D. Li, and W. Cheng, "A new steel defect detection algorithm based on deep learning," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–13, Mar. 2021, doi: 10.1155/2021/5592878.

[24] S. Wang, X. Xia, L. Ye, and B. Yang, "Automatic detection and classification of steel surface defect using deep convolutional neural networks," *Metals*, vol. 11, no. 3, p. 388, Feb. 2021, doi: 10.3390/met11030388.

[25] R. Tian and M. Jia, "DCC-CenterNet: A rapid detection method for steel surface defects," *Measurement*, vol. 187, Jan. 2022, Art. no. 110211, doi: 10.1016/j.measurement.2021.110211.

[26] M. Li, H. Wang, and Z. Wan, "Surface defect detection of steel strips based on improved YOLOv4," *Comput. Electr. Eng.*, vol. 102, Sep. 2022, Art. no. 108208, doi: 10.1016/j.compeleceng.2022.108208.

[27] L. Wang, X. Liu, J. Ma, W. Su, and H. Li, "Real-time steel surface defect detection with improved multi-scale YOLO-v5," *Processes*, vol. 11, no. 5, p. 1357, Apr. 2023, doi: 10.3390/pr11051357.

[28] S. Chaudhari, V. Mithal, G. Polatkan, and R. Ramanath, "An attentive survey of attention models," *ACM Trans. Intell. Syst. Technol.*, vol. 12, no. 5, pp. 1–32, Oct. 2021, doi: 10.1145/3465055.

[29] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020, doi: 10.1109/TPAMI.2019.2913372.

[30] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11531–11539.

[31] Z. Qin, P. Zhang, F. Wu, and X. Li, "FcaNet: Frequency channel attention networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 763–772.

[32] Y. L. Zhang, K. P. Li, K. Li, L. C. Wang, B. N. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Comput. Vis. (ECCV)*, Sep. 2018, pp. 286–301.

[33] K. Gregor, I. Danihelka, A. Graves, D. J. Rezende, and D. Wierstra, "DRAW: A recurrent neural network for image generation," in *Proc. Int. Conf. Mach. Learing*, 2015, pp. 1462–1471.

[34] J. Hu, L. Shen, S. Albanie, G. Sun, and A. Vedaldi, "Gather-excite: Exploiting feature context in convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2018, pp. 1–7.

[35] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7794–7803.

[36] Q.-L. Zhang and Y.-B. Yang, "SA-Net: Shuffle attention for deep convolutional neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 2235–2239.

[37] C.-Y. Wang, H.-Y. M. Liao, and I.-H. Yeh, "Designing network design strategies through gradient path analysis," 2022, *arXiv:2211.04800*.

[38] J. Chen, S.-h. Kao, H. He, W. Zhuo, S. Wen, C.-H. Lee, and S.-H. G. Chan, "Run, don't walk: Chasing higher FLOPS for faster neural networks," 2023, *arXiv:2303.03667*.

[39] K. Song and Y. Yan, "A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects," *Appl. Surf. Sci.*, vol. 285, pp. 858–864, Nov. 2013, doi: 10.1016/j.apsusc.2013.09.002.

[40] S. Xu, X. Wang, W. Lv, Q. Chang, C. Cui, K. Deng, G. Wang, Q. Dang, S. Wei, Y. Du, and B. Lai, "PP-YOLOE: An evolved version of YOLO," 2022, *arXiv:2203.16250*.

[41] C.-Y. Wang, A. Bochkovskiy, and H.-Y.-M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 7464–7475.

CHENGFEI LI is currently a Professor with Wuyi University. She has hosted and participated in multiple provincial-level projects. She has published over 20 articles in domestic and international journals. Her research interests include industrial control and fault detection, machine vision, artificial intelligence, and intelligent information processing.

AO XU graduated from Hubei University of Economics, Wuhan, China, in 2021. He is currently pursuing the master's degree with Wuyi University, Jiangmen, China. His research interest includes deep learning based defect detection.

QIBO ZHANG graduated from Nanyang Normal University, Nanyang, China, in 2021. He is currently pursuing the master's degree with Wuyi University, Jiangmen, China. His research interest includes deep learning based image processing.

YUFEI CAI graduated from Nanyang Institute of Technology, Nanyang, China, in 2022. He is currently pursuing the master's degree with Wuyi University, Jiangmen, China. His research interest includes deep learning based image processing.