International Journal of Engineering Inventions
e-ISSN: 2278-7461, p-ISSN: 2319-6491
Volume 14, Issue 6 [June. 2025] PP: 98-110

Steel defect detection algorithm based on ECBW-YOLO

Peiyong Ji[1], Ju Zhang[1], Dajian Fang[1], Jinbao Wu[1], Yutong Chen[1], Ying Huang[1], Ruiqiang Guo[1*]
[1]*College of Mechanical and Vehicular Engineering, Changchun University, 130022, Changchun, China*
[*]*Corresponding Author: Ruiqiang Guo*

ABSTRACT: *In response to the issues such as missed detections, false alarms, and difficulties in feature recognition for small - sized target defects in existing steel surface defect detection methods, this paper presents a steel defect detection algorithm grounded in the ECBW - YOLO model.Based on the YOLOv11s model, the Efficient Channel Attention (ECA) mechanism is incorporated into the C3K2 module, giving rise to a novel C3K2ECA module. This module can effectively capture the feature information of small - sized targets, thereby enhancing the feature extraction capabilities.Within the Neck network, the weighted Bidirectional Feature Pyramid Network (BIFPN) structure is introduced. This enables the model to detect targets across different scales and strengthens its detection performance for targets of varying sizes.Moreover, the Convolutional Block Attention Module (CBAM), which combines channel and spatial attention, is introduced. This module serves to suppress background interference and enhance the model's ability to capture global information. Additionally, the loss function Wise - IOU, which incorporates a dynamic non - monotonic focusing mechanism, is adopted to replace the original Complete Intersection over Union (CIOU) loss function, thus addressing the issue of bounding box overfitting.Experimental results demonstrate that, when using the NEU - DET dataset, compared with the YOLOv11s model, the mean Average Precision (mAP), the mean Average Precision at the IoU threshold range from 0.5 to 0.95 (mAP@0.5:0.95), and the Recall rate have increased by 1%, 1.3%, and 2.9% respectively.*

Keywords: *Steel, YOLOv11, C3k2ECA, BIFPN, CBAM, WIOU*

Date of Submission: 01-06-2025 Date of acceptance: 10-06-2025

I. INTRODUCTION

Steel ranks among the most widely used raw materials in the global manufacturing industry [1], playing a pivotal role in the development of modern civilization. Thanks to its outstanding properties, steel has become an essential material in construction, transportation, manufacturing, and mechanical applications. Notably, the surface quality of steel exerts a profound influence on its performance [2]. Therefore, ensuring the surface quality of steel is of utmost significance.

In the early days, the inspection of steel surfaces predominantly relied on manual methods [3]. However, due to the inherent limitations of human energy and attention, these methods suffered from low accuracy and inefficiency, consuming substantial amounts of time. As a result, they were unable to meet the requirements of contemporary steel surface inspection tasks.

In recent years, with the rapid advancement of deep learning, an increasing number of object detection methods based on deep learning have emerged. These methods can be broadly categorized into single - stage networks, such as the Single Shot MultiBox Detector (SSD) [4][5]and the You Only Look Once (YOLO) series [6][7][8]. There are also two - stage networks, including Region - CNN (R - CNN) [9], Towards Real - Time Object Detection with Region Proposal Networks (Faster R - CNN) [10], and Mask Region - based Convolutional Neural Network (Mask - CNN)[11]. Single - stage networks are characterized by high speed but relatively lower detection accuracy, while two - stage networks, although slower, offer higher detection precision. Among these, the YOLO series of algorithms strike a relatively good balance between detection accuracy and speed, making them particularly suitable for industrial defect detection.

Han J et al. [12] introduced the weighted bidirectional feature pyramid network structure (BIFPN) based on YOLOv5. This innovation optimized the fusion of feature maps, enabling the model to integrate information across multiple scales and thereby enhancing its ability to recognize and extract defects. Building on the work of Han J et al., Wang Y et al. [13] incorporated the channel attention mechanism (ECA) into the backbone of YOLOv7. This approach enhanced the importance of key feature channels, improving the algorithm's feature - learning capabilities and enabling it to capture more valuable information. Xin H et al. [14] further integrated the channel and spatial mixed attention module (CBAM) into YOLOv5, building on the research of Han J et al. This module can adaptively adjust feature maps by simultaneously considering spatial and channel features, thus providing a more comprehensive representation of the information within the feature maps. Shi J et al. [15] introduced the CBAM into YOLOv5, enhancing the model's focus on small - target defects and suppressing the interference of irrelevant information, which effectively improved the model's detection accuracy. Zhang L et al. [16] built upon the work of Shi J et al. by integrating the loss function with a dynamic non - monotonic focusing mechanism (WIOU), which mitigated the negative impacts of low - quality samples and geometric factors.

The main research contributions of this study are as follows:

1. The C3k2ECA module is introduced. By integrating the ECA attention mechanism into the C3k2 module within the Neck component, a minimal number of additional parameters are introduced. Through the interaction of information across different channels, this innovation enhances the algorithm's ability to extract and recognize the features of small - target defects, thereby improving the detection accuracy.

2. The BIFPN structure is incorporated. By fusing defect features at various scales, this addition strengthens the recognition of defect - related feature information, reducing the occurrences of missed detections and false alarms for small - target defects.

3. The CBAM spatial attention mechanism is adopted. By attenuating the influence of background elements in the dataset, this mechanism enhances the recognition of similar defects, further improving the detection accuracy of the algorithm. Additionally, the introduction of the WIoU loss function effectively addresses the issue of overfitting, thereby enhancing the training efficiency and stability of the algorithm.

II. YOLOv11s model

YOLOv11 is a YOLO - based detection model proposed in 2024. When compared with the YOLOv8 version, the network structure of YOLOv11 (as depicted in **Fig. 1**) has witnessed the following modifications:

The C3k2 mechanism has been put forward. In the shallow layers of the network, the c3k parameter is set to False, which bears resemblance to the C2f structure in YOLOv8. The C2PSA mechanism has been introduced. Specifically, a multi - head attention mechanism is embedded within the C2 mechanism. Depth - wise separable convolution (DWConv) has been incorporated. Two DWConvs are added to the classification and detection head. This addition serves to reduce the computational complexity and the number of parameters, thereby enhancing the efficiency of the model. An adaptive anchor box mechanism has been employed. This mechanism automatically optimizes the anchor box configuration for different datasets, thus improving the detection accuracy.

YOLOv11 is available in five versions, namely n, s, m, l, and x. In this study, the YOLOv11s version is chosen for further improvement.

As illustrated in **Fig. 1**:

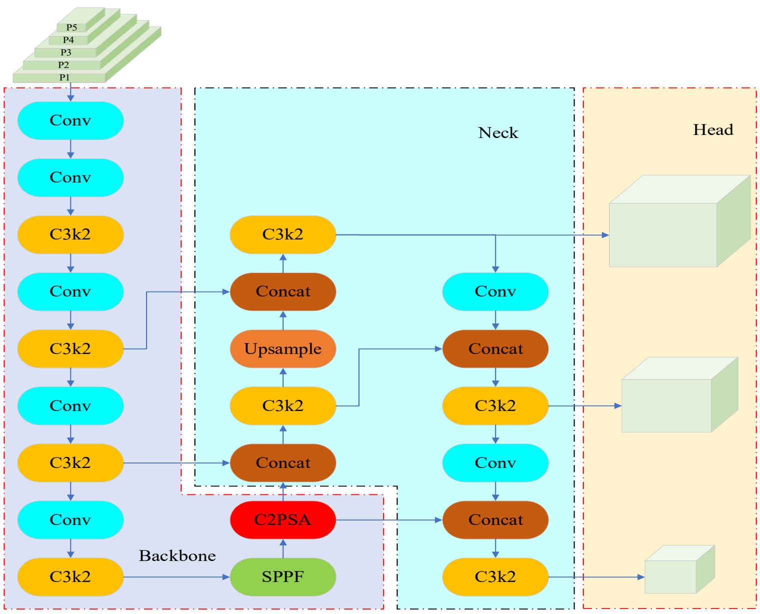

Fig. 1 YOLOv11s Model

III. The ECBW-YOLO model

The ECBW - YOLO model is developed by making improvements on the basis of the YOLOv11s model. Initially, the original C3K2 module is replaced by the C3K2ECA module. This substitution is designed to enhance the model's capacity to acquire features of complex and irregular targets.

Subsequently, BIFPN_Concat is employed to replace Concat in the neck component of the model. This modification aims to elevate the model's proficiency in dealing with defects of diverse sizes.

Furthermore, the CBAM attention mechanism is introduced into the neck region. This addition serves to strengthen the model's capabilities in feature recognition and aggregation.

Finally, the original CIOU is supplanted by WIOU, with the intention of improving the accuracy of the bounding boxes.

By means of these four improvements, the accuracy of the model in identifying defect types, sizes, and position variations is enhanced. This is graphically presented in **Fig. 2**.

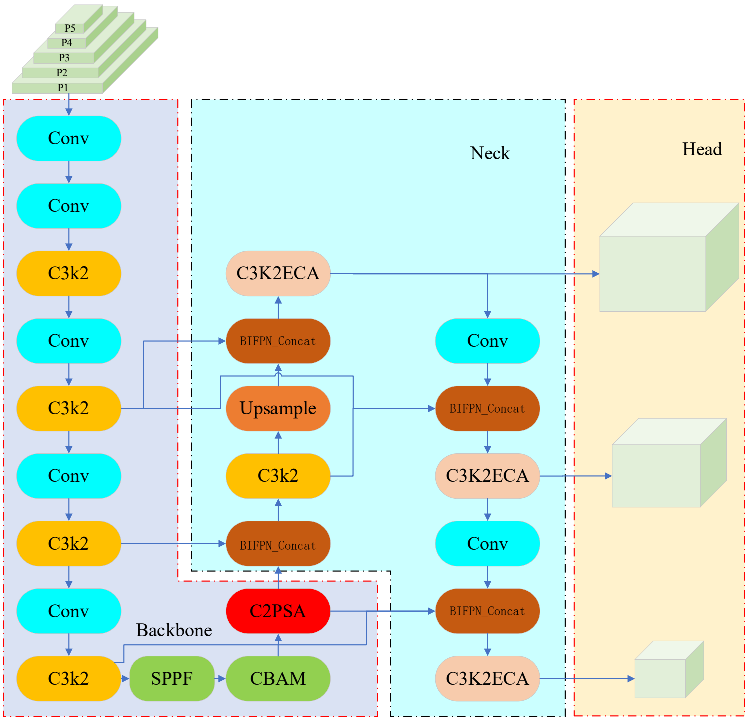

Fig. 2 the ECBW-YOLO Model

3.1 The newly improved C3ECA module

Cheng Z et al. [17] enhanced YOLOv5 by incorporating the ECA module to tackle the challenges associated with the identification of small and slender defects. In this study, the ECA module is employed to optimize the C3K2 module. The ECA module [18] belongs to the category of channel - level attention modules and represents an optimization of the SE module. It presents a novel approach that circumvents channel compression by facilitating local cross - channel interactions and adaptively determining the size of the one - dimensional convolutional kernel. This innovation leads to a notable improvement in performance. Specifically, the ECA module can effectively discern the interdependencies among channels, thereby significantly augmenting the feature extraction capabilities of convolutional layers.

The architecture of the ECA module is depicted in **Fig. 3**. Initially, the features of the original image are fed into the ECA module. Subsequently, global average pooling is carried out across all channels of the original image. Thereafter, a rapid one - dimensional convolution of size k is utilized to generate channel weights. These weights are then used to compute the corresponding probabilities for each channel. These probabilities are subsequently multiplied by the input features of the original image, and the resulting product serves as the input for the subsequent layer. The formula for k is：

$$K = \psi(C) = \left| \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{odd}$$

In the formula, K represents the size of the convolution kernel, C is the given channel dimension, and |t| odd is the odd number closest to t.

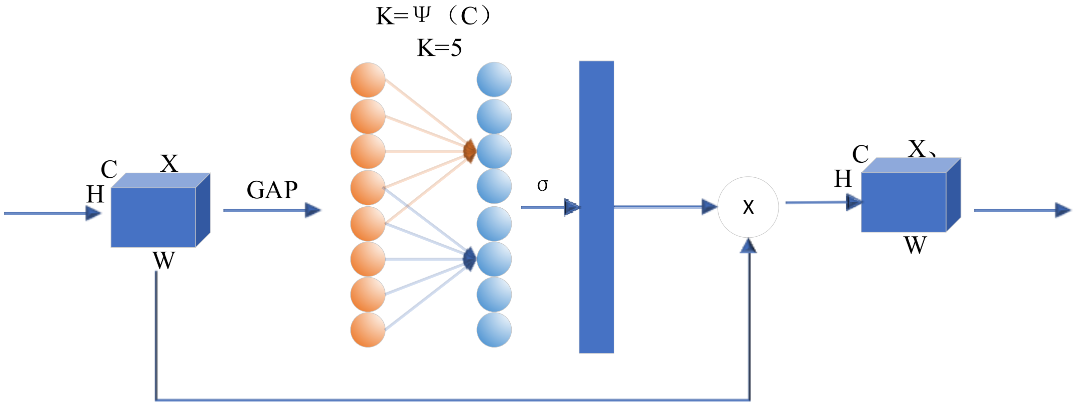

Fig. 3 ECA Module

The ECA module is integrated with the C3K2 module to yield a novel C3K2ECA module, as presented in **Fig. 4**.

In this study, by incorporating the C3K2ECA module, distinct weights are made to correspond to different convolutions. This approach serves to enhance the accuracy of recognition. Moreover, through the algorithmic interaction of information across various channels, the negative impact on model performance caused by dimensionality reduction is mitigated. This renders the algorithm relatively lightweight while concurrently maintaining the model's efficiency and computational efficacy.

Particularly, greater significance is attached to the feature information of small targets. As a result, the extraction of small - target features becomes more comprehensive.

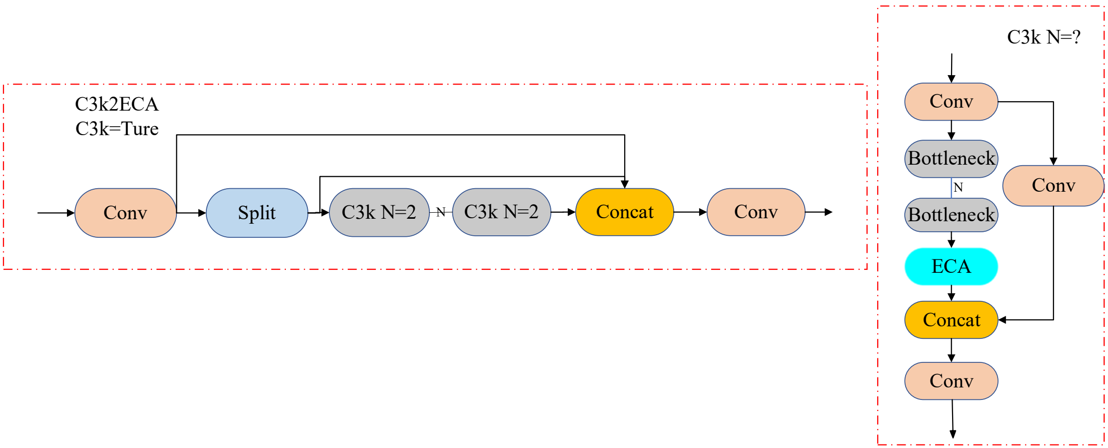

Fig. 4 C3k2ECA Module

3.2 Weighted Bidirectional Feature Pyramid Network (BIFPN)

Wu Y et al. [19] introduced the Bidirectional Feature Pyramid Network (BIFPN) to enhance the model's adaptability to targets at different scales, thereby significantly improving the detection accuracy. In this study, the BIFPN network is utilized to enhance the model's ability to recognize small-target defects.

As depicted in **Fig. 5**, the Feature Pyramid Network (FPN) fusion network aggregates features in a top-down manner. Nevertheless, this approach may result in a decline in the final accuracy because it fails to capture sufficient shallow-layer features. The fusion network employed in YOLOv11s is the Path Aggregation Network (PANet) structure, as presented in **Fig. 6**. This structure incorporates an additional bottom-up pathway on top of the FPN. Although this can address the limitations of the FPN, it increases the computational complexity.

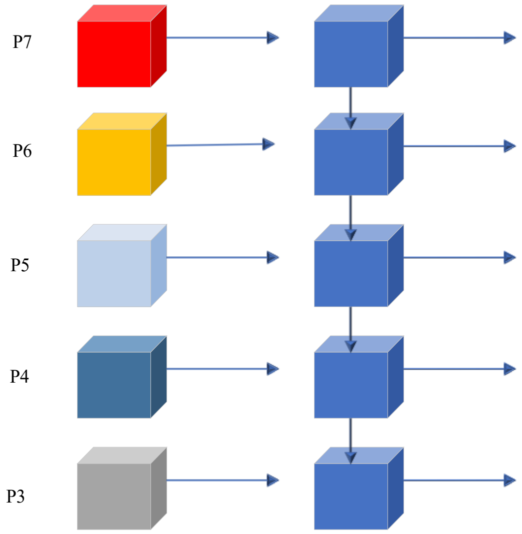

Fig. 5 FPN Structure

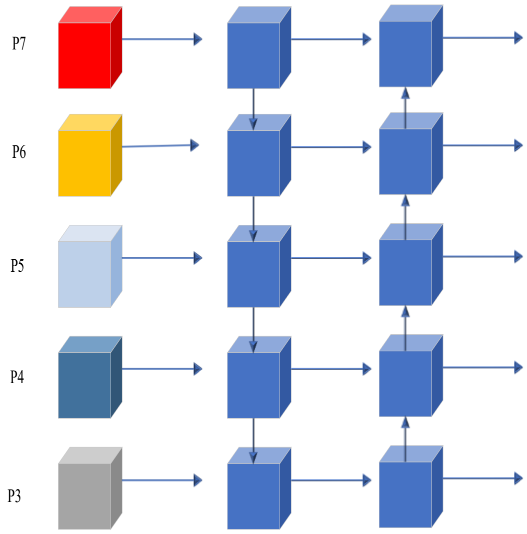

Fig. 6 Structure of PANet

The Weighted Bidirectional Feature Pyramid Network (BIFPN) is a feature extraction network optimized based on PANet. It employs both bottom - up and top - down approaches for feature aggregation. By introducing learnable weights, it is possible to enhance the detection accuracy.

As illustrated in **Fig. 7**, taking the P5 layer as an example, let $P5^{td}$ denote the intermediate feature of the P5 layer. The input feature is $P5^{in}$ and the output feature is $P5^{out}$. The corresponding formula is

$$P_5^{td} = Conv\left(\frac{\omega_1 \bullet P_6^{in} + \omega_2 \bullet \mathrm{Re}\,size\left(P_6^{in}\right)}{\omega_1 + \omega_2 + \varepsilon}\right) \tag{2}$$

$$P_5^{out} = Conv\left(\frac{\omega_1' \bullet P_5^{in} + \omega_2' \bullet P_5^{td} + \omega_3' \bullet \mathrm{Re}\,size\left(P_4^{out}\right)}{\omega_1' + \omega_2' + \omega_3' + \varepsilon}\right) \tag{3}$$

In the formula, ω represents the weights of each layer; Resize denotes the upsampling or downsampling operation; Conv represents the convolution operation; ε is a very small non - zero number.

Drawing on its principle, this study adds a channel to connect the input and output, integrating defect features at different scales. This approach aims to enhance the detection accuracy of surface defects on steel strips.

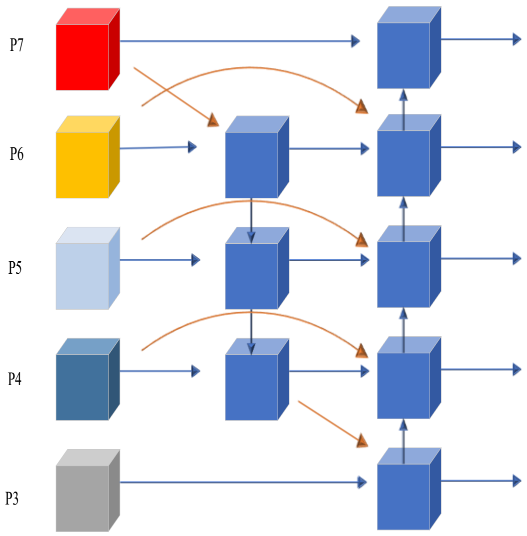

Fig. 7 BIFPN Structure

3.3 Channel and Spatial Mixed Attention Mechanism (CBAM)

The Convolutional Block Attention Module (CBAM) [20] incorporates a spatial attention mechanism on the foundation of the Squeeze-and-Excitation (SE) module, aiming to enhance the model's representational capacity.

The CBAM module encompasses a Channel Attention Module and a Spatial Attention Module. Its overall architecture is depicted in **Fig. 8**.

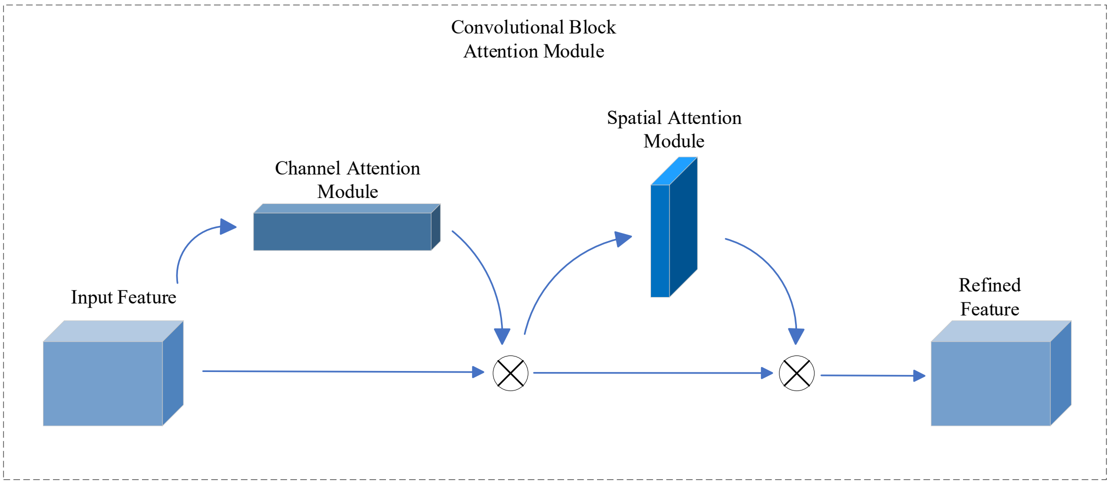

Fig. 8 CBAM Module

The channel attention mechanism can adaptively learn the weights of each channel to enhance the model's representational ability. The spatial attention module, on the other hand, focuses on the distinct feature representations within each channel to improve the model's accuracy.

The CBAM module infers a one - dimensional channel attention map $M_C \in R^{C\times1\times1}$ and a two - dimensional spatial attention map $M_S \in R^{1\times H\times W}$ from the input feature map $F \in R^{C\times H\times W}$. The process can be formulated as follows:

$$F' = M_C(F) \otimes F \tag{4}$$

$$F'' = M_S(F') \otimes F' \tag{5}$$

In the formula, $\otimes$ denotes element - wise multiplication, and F'' is the final extracted output.

The dataset employed in this study has a complex background. Moreover, some defects exhibit a high degree of similarity, which causes a decline in the accuracy of the detection algorithm. By introducing the CBAM module prior to upsampling, it is possible to mitigate background interference, enhance the recognition of defect features that resemble the background, more effectively capture global information features, and ultimately improve the performance of the model.

3.4 The loss function of the dynamic non-monotonic focusing mechanism (WIoU)

Li J et al. [21] employed the WIOU loss function to precisely measure the similarity among target frames, aiming to improve the defect recognition performance of the model. Han J et al. [22] introduced the WIOU loss function to tackle the problem of decreased accuracy resulting from uneven samples and to facilitate the accelerated convergence of the network. In this study, the WIoU loss function is adopted to replace the CIOU loss function.

The loss function for the network predicted bounding box coordinates of YOLOv11 is the CIOU loss, and its calculation formula is as follows:

$$L_{CIOU} = L_{IOU} + \frac{(\mathrm{x} - \mathrm{x}_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)} + \alpha v \tag{6}$$

$$\alpha = \frac{v}{L_{IOU} + v} \tag{7}$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w}{h} - \arctan\frac{w_{gt}}{h_{gt}}\right)^2 \tag{8}$$

$$L_{IOU} = \frac{W_i H_i}{wh + w_{gt}h_{gt} - W_i H_i} \tag{9}$$

In the formula, α represents the weighting function, whose role is to balance the parameters. v is the aspect ratio function, which is designed to ensure the uniformity of the height - width ratio. The representation of LIOU is presented in **Fig. 9**.

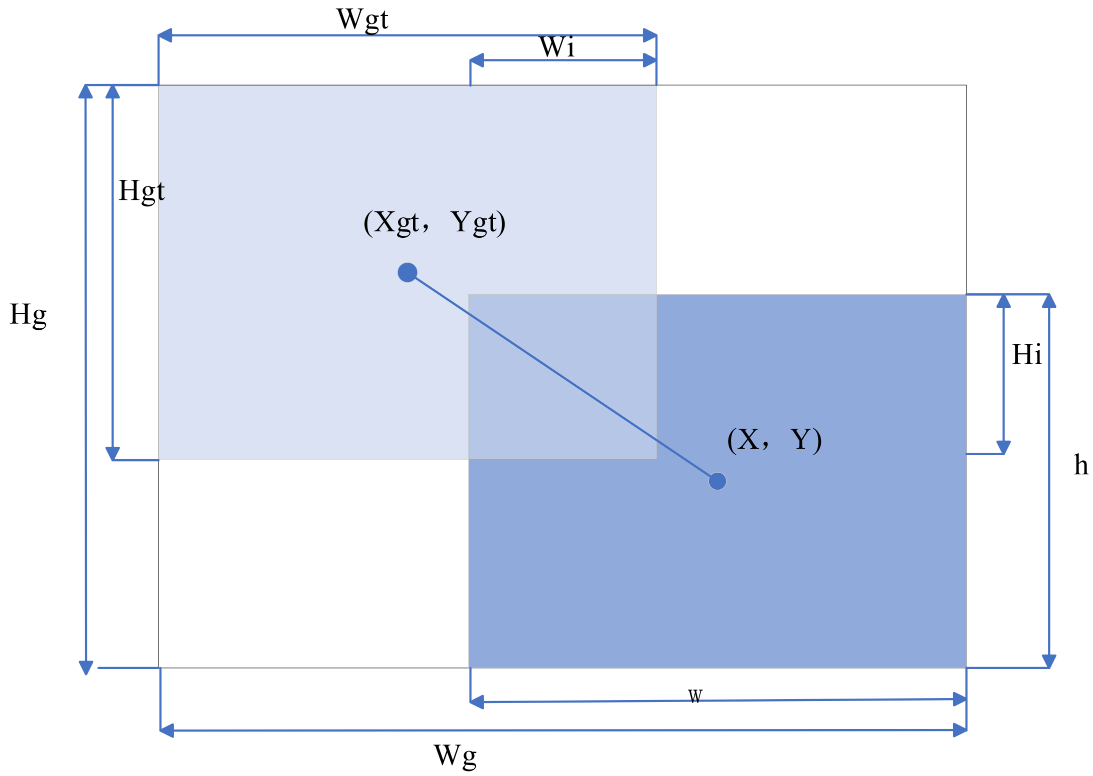

Fig. 9 LIOU Expression Diagram

Annotations: w, h, (X, Y) represent the width, height, and center coordinates of the predicted bounding box, respectively; W_{gt}, H_{gt}, (X_{gt}, Y_{gt}) represent the width, height, and center coordinates of the ground truth bounding box, respectively; W_i, H_i represent the width and height of the intersection area, respectively; W_g, H_g represent the width and height of the smallest enclosing box, respectively.

Based on the DIOU, the CIOU incorporates the loss related to the scale, length, and width of the detection bounding box. This enables the height of the predicted bounding box to match that of the actual one. Nevertheless, when the predicted bounding box and the actual bounding box exhibit a linear relationship, the penalty term of the CIOU degenerates to zero. Such a situation can be detrimental to the regression loss of the bounding box.

In this study, the WIOU loss function with a dynamic non - monotonic focusing mechanism is adopted to substitute the original CIOU. The formula of the WIOU loss function is presented as follows:

$$L_{WIOU} = rR_{WIOU}L_{IOU}, R_{WIOU} \in [1, e), L_{IOU} \in [0,1] \tag{10}$$

In the formula, the distance focusing mechanism RWIOU serves to magnify the LIOU of the moderately sized ordinary anchor boxes. The non - monotonic focusing coefficient r is employed to focus on the anchor boxes of ordinary quality. The defining equations of RWIOU and r are as follows:

$$R_{WIOU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \tag{11}$$

$$r = \frac{\beta}{\delta\alpha^{\beta-\delta}} \tag{12}$$

In the formula, δ represents the scaling constant that controls the weights of samples, and α is a hyperparameter denoting the focus coefficient. Through reducing the contribution of high - quality samples to the loss value, r dynamically adjusts the gradient gain of the bounding box. Additionally, in the later stage of training, it mitigates harmful gradients, focuses on ordinary anchor boxes, and enhances the positioning ability.

The outlier degree is defined to characterize the quality of anchor boxes, and its definition is as follows.

$$\beta = \frac{\gamma_{IOU}^*}{\overline{\gamma_{IOU}}} \in [0, +\infty) \tag{13}$$

In the formula, $\overline{\gamma_{IOU}}$ represents the dynamic moving average value. Its function is to assign a relatively small gradient gain when the value is either large or small, thereby reducing the impact on the bounding box. γ_{IOU}^* denotes the threshold between the predicted bounding box and the ground truth bounding box. The smaller the value of β, the higher the quality of the anchor box; conversely, the lower the quality.

The WIOU loss function balances the impacts of high - quality and low - quality anchor boxes on the model. It enhances the model's generalization ability and optimizes the model's performance. In this paper, the WIOU loss function is employed to substitute the original CIOU loss function for model optimization.

IV. Experimental Results and Analysis

4.1 Evaluation criteria

Regarding the criteria for evaluating the steel surface inspection model, the commonly adopted ones are as follows: Average Precision (AP), Mean Average Precision (mAP), Frames Per Second (fps), Number of Parameters, Precision, and Recall. Herein, AP represents the area under the Precision - Recall (PR) curve, which is utilized to assess the detection accuracy of each type of defect. mAP is the mean value of the AP of all types, serving to evaluate the detection accuracy of all defects. The expression is given by:

$$AP_{\mathrm{i}} = \int_0^1 P(R)dR \tag{14}$$

$$\mathrm{m}AP = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{n} \mathrm{AP_i} \tag{15}$$

In the formula, P denotes the precision and R represents the recall. The corresponding formulas are given as follows:

$$P = \frac{T_P}{T_P + F_P} \tag{16}$$

$$R = \frac{T_P}{T_P + F_N} \tag{17}$$

In the formula, T_P refers to the number of targets that are correctly detected, F_P represents the number of targets that are detected incorrectly, and F_N stands for the number of targets that are missed during the detection process.

4.2 Experimental environment and dataset

The experimental environment employed in the course of this experimental study is as follows: The operating system is Windows 10; the GPU is NVIDIA GeForce GTX 1650; the compilation environment consists of Python 3.8.18, PyTorch 1.10.0, and CUDA 10.2. During the model training process, the dataset was partitioned into a training set, a validation set, and a test set at a ratio of 8:1:1. The experimental parameter settings are presented in **Table 1** below.

Table 1 Experimental Parameter Settings

Name	Parameter settings
Image size	224
Batch size	8
Epochs	300
Workers	4
close-mosaic	10

The dataset adopted in this experiment is the publicly released NEU - DET dataset by Northeastern University. This dataset encompasses six defect characteristics, namely rolled - in scale (RS), crazing (Cr), pitted surface (PS), patches (Pa), scratches (Sc), and inclusions (In). Each type of defect has 300 images, resulting in a total of 1800 images in the dataset. To enhance the robustness of the images, operations such as brightening and darkening were carried out on the dataset. The results of the quantity processing for each type of defect are presented in **Table 2** as follows. For the purpose of effective model training, the dataset was partitioned into a training set, a validation set, and a test set at a ratio of 8:1:1.

Table 2 Dataset Processing

Defects	Original	Brightening and darkening processing
rolled-in_scale(RS)	300	900
crazing(Cr)	300	900
pitted_surface (PS)	300	900
patches(Pa)	300	900
scratches(Sc)	300	900
inclusion(In)	300	900
All	1800	5400

4.3 Melting experiment

To validate the effectiveness of the improvements put forward in this paper, ablation experiments were devised on the NEU - DET dataset, with YOLOv11s serving as the foundation. To guarantee the consistency of parameters throughout the experimental process, no pre - trained weights were configured.

In the table, "a" represents the utilization of the YOLOv11s module; "b" denotes the incorporation of the C3K2ECA module into the YOLOv11s module; "c" signifies the addition of both the C3K2CEA module and the BIFPN structure to the YOLOv11s module; "d" indicates the introduction of the C3K2ECA module, the BIFPN structure, and the WIOU loss function to the YOLOv11s module; and "e" stands for the module designed in this paper. The experimental results are presented in **Table 3** as follows:

Table 3 Ablation Experiments

Class	AP						Precision	mAP	mAP@0.5:0.95	recall
	RS	Cr	PS	Pa	Sc	In				
a	96.0	92.3	96.2	99.3	98.3	97.4	93.6	96.6	67.0	91.9
b	96.3	93.8	96.5	99.2	98.9	97.2	93.8	97.0	67.0	93.6
c	97.2	92.8	96.0	99.2	98.9	97.5	93.8	96.9	67.3	93.8
d	97.5	95.0	96.6	99.2	99.0	97.4	93.5	97.4	67.9	94.4
e	97.4	94.8	96.8	99.4	99.0	98.0	93.6	97.6	68.3	94.8

The result analysis reveals the following: In comparison to the experiments in Group a, in the experiments of Group b, Precision increased by 0.2%, mAP (mean Average Precision) increased by 0.4%, mAP@0.5:0.95 remained unchanged (an increase of 0%), and recall rose by 1.7%.

For Group c experiments relative to those of Group a, Precision was enhanced by 0.2%, mAP increased by 0.3%, mAP@0.5:0.95 increased by 0.3%, and recall improved by 1.9%.

Regarding Group d experiments as compared to Group a, Precision decreased by 0.1% (an increase of - 0.1%), mAP increased by 0.8%, mAP@0.5:0.95 increased by 0.9%, and recall increased by 2.5%.

Finally, in the case of Group e experiments in contrast to Group a, Precision showed no change (an increase of 0%), mAP increased by 1%, mAP@0.5:0.95 increased by 1.3%, and recall increased by 2.9%.

4.4 Comparative test

To verify the effectiveness and superiority of this algorithm in steel material detection, the dataset was brightened and darkened respectively, and then compared with other mainstream detection methods as shown in **Table 4** below:

Table 4 Comparative Experiment

Class	mAP	Precision	mAP@0.5:0.95	recall	Model size/M	GFLOPs/G
SSD	93.5	91.1	69.5	85.1	95.3	61.1
RT-DETR	96.2	93.0	74.2	91.9	16.2	103.5
YOLOv3	95.0	92.4	68.4	91.2	204.8	283.0
YOLOv5s	94.5	92.2	65.0	91.9	13.9	15.8
YOLOv6	94.8	92.4	71.1	90.9	32.0	44.0
YOLOv8s	95.2	93.1	71.3	90.3	21.9	28.4
YOLOv11s	96.6	93.6	67.0	91.9	19.2	21.3
ECBW-YOLO	97.6	93.6	68.3	94.8	24.2	26.3

Based on the analysis of the experimental results, when compared with the SSD, RT - DETR, YOLOv3, YOLOv5s, YOLOv6, YOLOv8s, and YOLOv11s models, the mAP of the algorithm proposed in this study has witnessed increases of 4.1%, 1.4%, 2.6%, 3.1%, 2.8%, 2.4%, and 1% respectively. The Precision has been enhanced by 2.5%, 0.6%, 1.2%, 1.4%, 1.2%, 0.5%, and 0% respectively. The recall has risen by 9.7%, 2.9%, 3.6%, 2.9%, 3.9%, 4.5%, and 2.9% respectively.

In light of the complexity of the models, it can be observed that the model developed in this paper, which is an improvement upon YOLOv11, has significantly fewer parameters, a smaller model size, and lower GFLOP values compared to current state - of - the - art YOLO algorithms such as YOLOv8. Thus, the model presented herein demonstrates excellent detection performance despite having fewer parameters compared to existing advanced algorithms, rendering it more suitable for general industrial inspection scenarios.

4.5 Visual qualitative analysis

The algorithm employed in this study was compared with YOLOv11s using the dataset that had undergone brightening and darkening treatments. The outcomes are presented in **Fig. 10** and **Fig. 11**. The area bounded by the curve, the vertical axis, and the horizontal axis corresponds to the Average Precision (AP) value of the respective category. The nearer the curve approaches the upper - right corner, the more favorable the performance, and the greater the AP value. The blue curve denotes the mean Average Precision (mAP) of the six types of defects, while the other colored curves represent the AP values of other individual defects. The AP values of each defect type exhibited variations, and the mAP value was optimized.

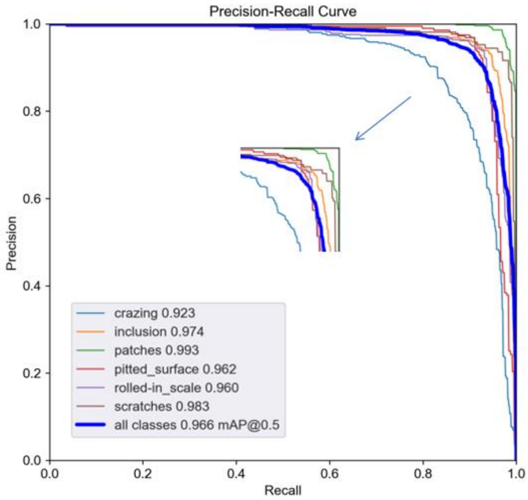

Fig. 10 PR Curve of YOLOv11s

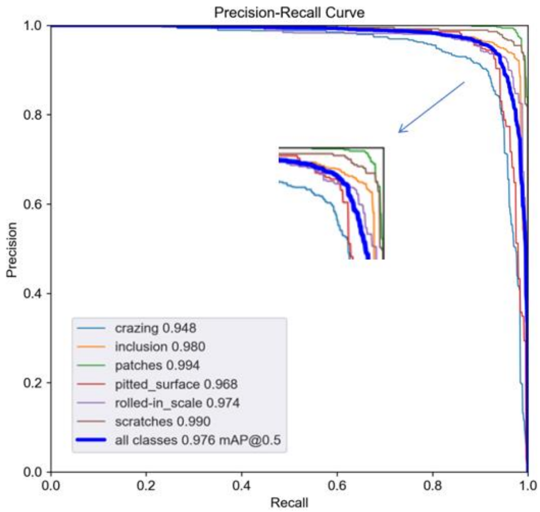

Fig. 11 PR Curve of ECBW-YOLO

As depicted in **Fig. 12**, a comparison is made between the approach presented in this study and other mainstream algorithms. Evidently, the proposed method has effectively decreased the missed detection rate of small targets, leading to a notable enhancement in accuracy.

Specifically, the refined model has alleviated the issue of missed detections in steel defect inspections, achieving more precise positioning. This improvement not only reflects the effectiveness of the model optimization but also demonstrates its potential application value in practical steel defect detection scenarios.

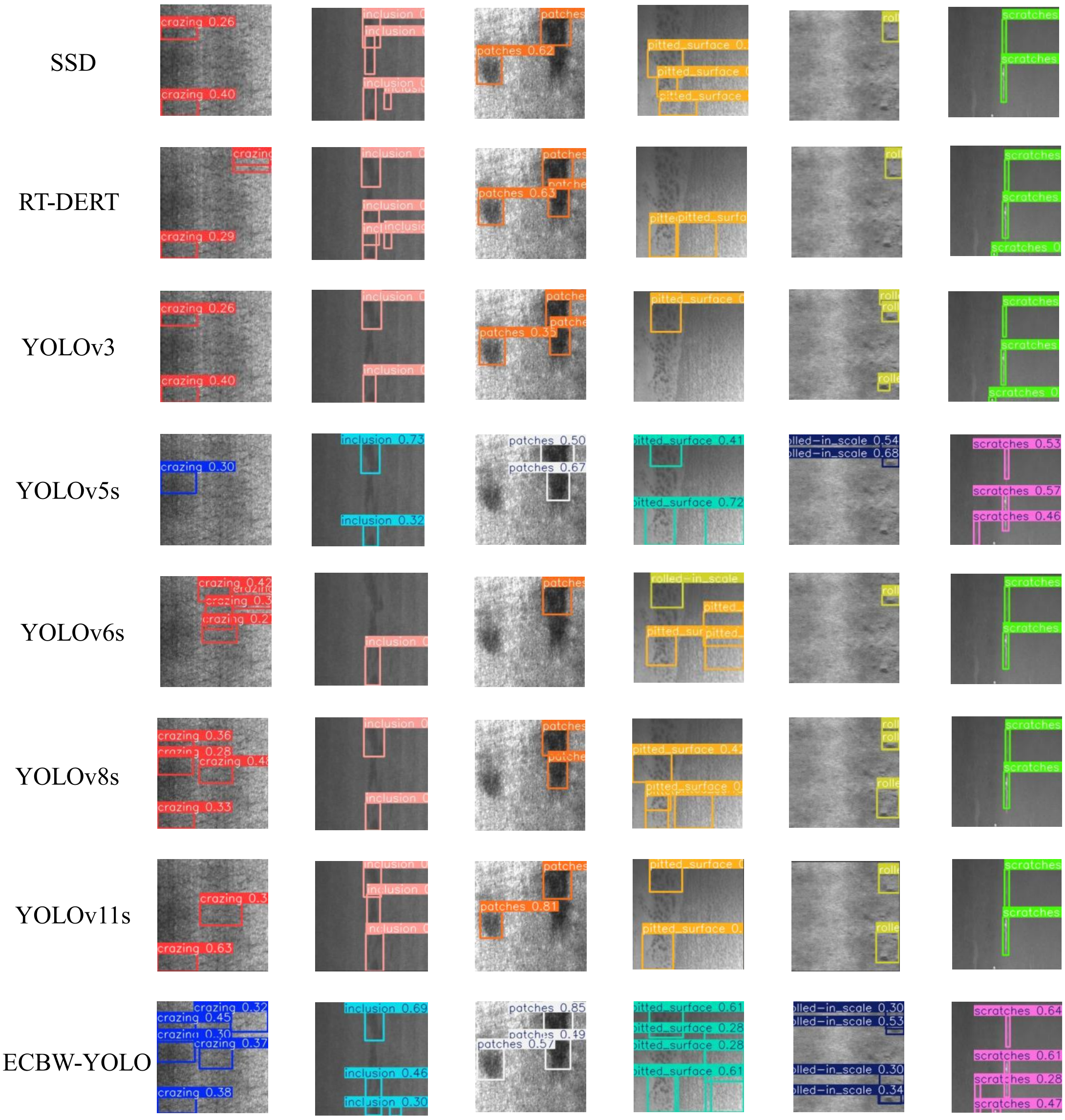

Fig. 12 Comparison of Effects

V. Conclusion

This study presents a steel defect detection algorithm grounded in the ECBW - YOLO model. Aiming at the prevalent issues in steel defect detection, such as missed detections of small targets and suboptimal detection quality, modifications have been made to the YOLOv11s model.

Specifically, the C3K2ECA module and the BIFPN network structure are incorporated. This integration facilitates the algorithm to more effectively fuse information features, thereby enhancing its ability to precisely locate small targets and concomitantly reducing the missed detection rate of such targets. The introduction of the CBAM attention mechanism module serves to accentuate the feature information of target defects while attenuating the influence of background factors, thus enabling more accurate localization of the feature information of defect targets. Additionally, the WIOU loss function is adopted to supplant the CIOU loss function, which not only improves the detection accuracy but also addresses the problem of overfitting in boundary detection boxes.

Experimental findings indicate that, in comparison with existing models, the model refined from YOLOv11s in this research demonstrates a notable improvement in both mAP (mean Average Precision) and Recall. Moreover, it successfully mitigates problems such as missed detections and low detection accuracy for small - target defects. The model also exhibits enhanced lightweight characteristics in terms of model size and parameter count, thereby meeting the demands of industrial production.

Notwithstanding these achievements, there remain challenges in achieving high - precision detection for certain defects, such as cracks and pitting. The data employed in this study is sourced from a public dataset, and it is well - recognized that the improvement of detection accuracy is significantly influenced by the data collection environment. In real - world industrial applications, numerous factors can impact the detection accuracy of the model. These factors include light intensity, lighting angle, humidity, and noise, among others, to which the detection accuracy is particularly sensitive.

Looking ahead, our future endeavors will involve collecting a more extensive range of data samples encompassing the aforementioned factors for model training. While maintaining a focus on detection accuracy, we will also strive to enhance the model's lightweight properties and detection speed, rendering it more conducive to industrial applications. The primary areas of focus for future research are as follows:

In the realm of steel surface defect detection, the process of defect annotation for images typically requires the expertise of professionals, which incurs substantial labor costs. Moreover, the quantity of data samples significantly influences the performance of the model. Consequently, in the future, we aim to leverage unsupervised learning techniques to substitute supervised learning, thereby alleviating the challenges and costs associated with data generation.

Regarding the detection model, we intend to introduce more efficient detection heads. These heads will endow the model with the capacity to recognize information features based on the size, shape, and position of the targets, thereby enhancing detection accuracy. Additionally, we will incorporate more lightweight convolutional operations to further streamline the model while effectively extracting feature information across diverse scales, thus improving the overall detection performance. Furthermore, we will explore the integration of superior attention mechanisms that can autonomously adjust the mechanism weights in accordance with the characteristics of target defects, thereby facilitating more effective target detection.

REFRENCES

[1]. He K, Wang L. A review of energy use and energy-efficient technologies for the iron and steel industry[J]. Renewable and Sustainable Energy Reviews, 2017, 70: 1022-1039.

[2]. Rudawska A. Selected aspects of the effect of mechanical treatment on surface roughness and adhesive joint strength of steel sheets[J]. International journal of adhesion and adhesives, 2014, 50: 235-243.

[3]. Gao Y, Gao L, Li X, et al. A semi-supervised convolutional neural network-based method for steel surface defect recognition[J]. Robotics and Computer-Integrated Manufacturing, 2020, 61: 101825.

[4]. Zhai S, Shang D, Wang S, et al. DF-SSD: An improved SSD object detection algorithm based on DenseNet and feature fusion[J]. IEEE access, 2020, 8: 24344-24357.

[5]. Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.

[6]. Kou X, Liu S, Cheng K, et al. Development of a YOLO-V3-based model for detecting defects on steel strip surface[J]. Measurement, 2021, 182: 109454.

[7]. Guo Z, Wang C, Yang G, et al. Msft-yolo: Improved yolov5 based on transformer for detecting defects of steel surface[J]. Sensors, 2022, 22(9): 3467.

[8]. Xiao D, Xie F T, Gao Y, et al. A detection method of spangle defects on zinc-coated steel surfaces based on improved YOLO-v5[J]. The International Journal of Advanced Manufacturing Technology, 2023, 128(1-2): 937-951.

[9]. Cai Z, Vasconcelos N. Cascade r-cnn: Delving into high quality object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6154-6162.

[10]. Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 39(6): 1137-1149.

[11]. Dollár K H G G P, Girshick R. Mask r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.

[12]. Han J, Cui G, Li Z, et al. DBCW-YOLO: A Modified YOLOv5 for the Detection of Steel Surface Defects[J]. Applied Sciences, 2024, 14(11): 4594.

[13]. Wang Y, Wang H, Xin Z. Efficient detection model of steel strip surface defects based on YOLO-V7[J]. Ieee Access, 2022, 10: 133936-133944.

[14]. Xin H, Zhang K. Surface defect detection with channel-spatial attention modules and bi-directional feature pyramid[J]. IEEE Access, 2023.

[15]. Shi J, Yang J, Zhang Y. Research on steel surface defect detection based on YOLOv5 with attention mechanism[J]. Electronics, 2022, 11(22): 3735.

[16]. Zhang L, Fu Z, Guo H, et al. Multiscale local and global feature fusion for the detection of steel surface defects[J]. Electronics, 2023, 12(14): 3090.

[17]. Cheng Z, Gao L, Wang Y, et al. EC-YOLO: Effectual Detection Model for Steel Strip Surface Defects Based on YOLO-V5[J]. IEEE Access, 2024.

[18]. Liming Z ,Shixin L ,Zhiren Z , et al.Improved YOLOv5 foreign object detection for transmission lines[J/OL].Optoelectronics Letters,1-7[2024-05-31].http://kns.cnki.net/kcms/detail/12.1370.TN.20240423.1454.006.html.

[19]. Wu Y, Chen R, Li Z, et al. SDD-YOLO: A Lightweight, High-Generalization Methodology for Real-Time Detection of Strip Surface Defects[J]. Metals, 2024, 14(6): 650.

[20]. Lu J, Zhu M, Ma X, et al. Steel Strip Surface Defect Detection Method Based on Improved YOLOv5s[J]. Biomimetics, 2024, 9(1): 28.

[21]. Li J, Chen M. DEW-YOLO: An Efficient Algorithm for Steel Surface Defect Detection[J]. Applied Sciences, 2024, 14(12): 5171.

[22]. Han J, Cui G, Li Z, et al. DBCW-YOLO: A Modified YOLOv5 for the Detection of Steel Surface Defects[J]. Applied Sciences, 2024, 14(11): 4594.