

## Research Article

# The Steel Surface Multiple Defect Detection and Size Measurement System Based on Improved YOLOv5

Yiming Xu,<sup>1</sup> Ziheng Ding ,<sup>1</sup> Wang Li,<sup>1</sup> Kai Zhang ,<sup>1</sup> and Le Tong <sup>2</sup>

<sup>1</sup>School of Electrical Engineering, Nantong University, Nantong 226019, China

<sup>2</sup>Shanghai Normal University, Shanghai 200234, China

Correspondence should be addressed to Le Tong; [tongle@shnu.edu.cn](mailto:tongle@shnu.edu.cn)

Received 16 March 2023; Revised 28 April 2023; Accepted 2 May 2023; Published 20 May 2023

Academic Editor: Yang Li

Copyright © 2023 Yiming Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the process of steel production, the defects on the surface of steel will adversely affect the subsequent processing of a product. Accurate detection of such defects is the key to improve production efficiency and economic benefits. In this paper, an end-to-end steel surface defect detection and size measurement system based on the YOLOv5 model is designed. Firstly, in consideration of the defect location and direction correlation in the production process, a coordinate attention mechanism is added at the head of YOLOv5 to strengthen the spatial correlation of the steel surface and an adaptive anchor box generation method based on defect shape difference feature is proposed, which realizes the detection of three main types of defects on the Pytorch deep learning framework. Secondly, BiFPN is used to strengthen the feature fusion and a transformer encoder is added to improve the performance of detecting small defects. Thirdly, calculate the conversion ratio between the pixel and the actual size according to the standard reference specimen and obtain the actual size through the pixel statistics of the defect area to achieve pixel level size measurement. Finally, the steel surface defect detection and size measurement system are designed in this paper, which consist of various hardware, related measurement, and detection algorithms. According to the experimental results, the comprehensive defect detection accuracy of this method reaches 93.6%, of which the scratch detection accuracy reaches 95.7%. The detection speed reaches 133 fps and the defect size measurement accuracy reaches 0.5 mm. Experimental result shows that the defect detection and size measurement system designed in this paper can accurately detect and measure various industrial production defects and can be applied to the actual production process.

## 1. Introduction

Surface defects detection is one of the most important factors affecting the quality of steel products. Some defects on the steel surface will not only affect the subsequent production, but also affect the corrosion resistance and wear resistance of the final product. Therefore, how to accurately detect steel surface defects has become an important topic in the field of steel production. In recent years, machine vision-based detection methods are known for their advantages of noncontact and fast response. The detection method is very important for the detection system, which not only determines the accuracy and real-time performance of defect identification and classification, but also guarantees the subsequent tasks, such as defect location and defect size measurement.

The defect detection based on machine vision is a non-contact detection; without manual intervention, just select the appropriate camera and light source to collect the product surface image and then use the relevant defect detection algorithm to identify and classify defects. This method has high detection efficiency and can greatly improve the level of manufacturing industry. The defect detection based on machine vision can be divided into traditional methods and deep learning methods. The defect detection method based on traditional machine learning is mainly based on texture, color, and shape features. Deng et al. [1] proposed an improved RAPID (reconstruction algorithm for probabilistic inspection of defect) imaging method for detecting defects in composite plates using machine learning-based feature identification. Wu et al. [2] established a rail defect detection model based on the

color and shape features obtained from the ultrasonic scanning of the train track and the characteristics of the rail defect area distribution and converted the detection problem to the contour classification problem and improved the detection performance of the three kinds of defects. Liu et al. [3] used an improved local binary mode algorithm to extract the texture features of steel plates on the production line to achieve real-time defect detection under high temperature environment. Czimmermann et al. [4] by optimizing Gabor filters and using machine learning methods based on statistics and structure for texture defect detection can not only detect steel surface defects, but also overcome the problem that deep learning requires a lot of computing resources. Sun et al. [5] used the MBS method to extract the features of the weld defect and the classification algorithm based on the Gaussian mixture model, which can not only effectively identify the weld defects, but also accurately distinguish the weld defects types. The abovementioned traditional defect detection methods need to manually determine the features. For complex scenes, the manually extracted features cannot fully express the feature information, which has certain limitations.

In recent years, due to the significant improvement of computer computing power, defect detection using deep learning is more accurate and efficient than traditional methods. Park et al. [6], based on the YOLOv3 tiny model, detected the cracks on the surface of concrete buildings and calculated the size of cracks with laser beam and distance sensor, achieving real-time effects. Zhao et al. [7] proposed a dual channel faster R-CNN [8], which divides defect location and recognition into two network branches and combines the idea of superresolution enhancement to detect defects in key parts of the railway. Wu et al. [9] proposed an end-to-end learning method for industrial defect detection, achieving high accuracy and reducing the dependence on image resolution. Guan et al. [10] proposed a defect detection method for specular surfaces using a combination of deflectometry and deep learning techniques, which solved the shortcoming that specular reflection leads to unclear features. Kou et al. [11] developed an end-to-end defect detection model based on YOLOv3. The proposed model selected the ideal feature scale by removing the anchor frame and enhanced the network's characterization capability with dense convolutional blocks. He et al. [12] proposed a convolutional neural network incorporating VGG16 [13] to improve the classification performance and input the classification results into YOLO to obtain better detection efficiency. Ihor et al. [14] put forward a recognition classifier for scratch and wear on metal surfaces based on ResNet50 and ResNet152 deep residual neural network architecture [15], which improved the accuracy of steel surface defect detection. Sharma et al. [16] combined with the Unet [17] semantic segmentation model to detect scratches on the steel surface and achieved good accuracy. Xie et al. [18] reduced the YOLO model by combining deep separable convolution and established a new prediction bounding box loss function to accelerate the convergence of the model. Zhang et al. [19] improved the PP-YOLOE-m strip surface defect identification network by using the anchor-free method without considering the optimization of the anchor box, reducing the

superparameters, and improving the accuracy of strip surface defect detection. Guan et al. [20] evaluated the quality of feature image through decision tree and adjusted the parameters and structure of VGG19 to obtain a new target detection network, which significantly improved the detection accuracy of steel surface defects. Yeung and Lam [21] proposed a fused-attention network combined with adaptively balanced feature fusion algorithm for steel surface defect detection, which improved the accuracy and speed of steel surface defect detection. Liu et al. [22] designed the periodic rolling mark defect generator, proposed the CNN + LSTM [23] detection algorithm, and improved the detection method based on the attention mechanism algorithm, thus improving the rolling mark defect recognition ability. Damacharla et al. [24] proposed a U-net framework based on transfer learning to detect steel surface defects, which demonstrated that the convergence speed of transfer learning is better than random initialization, but for rare defect types and complex shape defects, the performance of transfer learning is poor. Yang et al. [25] proposed an FDIA detection method based on secure federated deep learning, using transformers to study the relationship between individual electricity quantities and using a joint learning framework to collaboratively train the detection model. While improving detection efficiency, it also protects data privacy and reduces communication overhead. While deep learning-based defect detection methods have been extensively studied, the challenge of poor detection and classification performance resulting from variations in the shape of steel surface defects remains unresolved. In this paper, we address this issue through related research efforts.

The steel surface defect detection and size measurement system based on machine vision proposed in this paper realizes defect detection through the improved YOLOv5 deep learning model and completes the real size measurement of defects through the Canny edge detection algorithm and pixel size converted by CCD camera internal parameters. This paper combines the self-built dataset and NEU-DET open source dataset, by adding coordinate attention [26] mechanism and improving the clustering method of the YOLOv5 adaptive anchor frame, it realizes accurate detection of scratches, inclusion, patches, and other types of defects and can measure the defect size of actual objects. Its detection accuracy and measurement accuracy meet the requirements of practical engineering application scenarios.

## 2. Related Works

*2.1. Object Detection.* YOLOv5 is an object detection model based on deep learning. The model can achieve high reasoning speed under the premise of ensuring certain accuracy. Therefore, the YOLOv5 model is selected for the test.

*2.1.1. YOLOv5 Model.* YOLOv5 essentially realizes a deep convolutional network with regression function. As the initial step in the deep convolutional network, image feature information is extracted. In contrast to two-stage detection

algorithms that generate candidate boxes and then extract features, the YOLOv5 algorithm conducts end-to-end training and reasoning for all regions of the entire image, resulting in faster processing speeds and superior foreground-background distinction. In addition, the model fuses features of different scales, which makes this algorithm more conducive to identifying small targets and suitable for defect detection. The second step of objects detection is to classify the extracted features. The model divides the whole picture into multiple regions. If the center of the object to be detected falls in this region, the classification model is used to classify this region.

For each network, there is a bounding box. Each prediction will generate four parameters to determine the coordinates of the bounding box, namely, the coordinates  $x$  and  $y$  of the upper left corner, the width and the height. Confidence will also be generated according to the logistic

regression. Confidence will be used to judge whether the bounding box contains targets. If the confidence degree is high, which means the bounding box contains targets, the next step of target classification will be performed. YOLOv5 network structure is shown in Figure 1. And the steel surface detect model proposed in this paper is shown in Figure 2.

### 2.1.2. Loss Function and Generation of Adaptive Anchor Box.

Loss function is an important part of deep learning. In the process of iterative learning, deep learning is to continuously reduce the value of loss function.

The loss function of the YOLOv5 model consists of three parts, namely, the position offset loss, confidence loss, and classification loss of the bounding box. The loss function is shown in the following equation:

$$L_{v5}(t_p, t_{gt}) = \sum_{k=0}^K \left[ \alpha_k^{\text{balance}} \alpha_{\text{box}} \sum_{i=0}^{S^2} \sum_{j=0}^B \prod_{kij}^{\text{obj}} L_{\text{obj}} + \alpha_{\text{obj}} \sum_{i=0}^{S^2} \sum_{j=0}^B \prod_{kij}^{\text{obj}} L_{\text{obj}} + \alpha_{\text{cls}} \sum_{i=0}^{S^2} \sum_{j=0}^B \prod_{kij}^{\text{obj}} L_{\text{cls}} \right] \quad (1)$$

where  $K$ ,  $S^2$ , and  $B$  are the output feature map, bounding box, and the number of anchors on each bounding box, respectively.  $\alpha_{\text{box}}$  is the weight of the loss of the bounding box,  $\alpha_{\text{cls}}$  is the weight of the loss of the classification, and  $\alpha_{\text{obj}}$  is the weight of the loss of the confidence.  $\prod_{kij}^{\text{obj}}$  indicates whether the  $k$ -th output feature map, the  $i$ -th prediction box, and the  $j$ -th anchor box is positive samples. If it is the positive sample,  $\prod_{kij}^{\text{obj}}$  equals to 1, and if not, it is 0;  $t_p$  is the prediction vector and  $t_{gt}$  is the ground truth vector.  $\alpha_k^{\text{balance}}$  is used to balance the weight of the output feature map of each scale. The default value is (4.0, 1.0, and 0.4), corresponding to the output feature map of  $80 \times 80$ ,  $40 \times 40$ ,  $20 \times 20$  successively. The learning process of YOLOv5 is to continuously reduce the loss function, that is, the predicted value and the real value are getting closer and closer, so as to achieve better detection accuracy.

The series of YOLO algorithms will generate a series of anchor boxes with fixed positions before training data. These anchor boxes can be regarded as candidate regions where the target is located. Selecting an appropriate anchor box will make it easier for the network model to obtain a good detector. For the steel surface defect detection in this paper, the shapes of different kinds of defects on the steel surface vary greatly. For example, the shapes of scratches and inclusion are long strips, with large differences in length and width, while the patches are almost elliptical. Also, even the same kind of defects will have large differences in shape. If the preset anchor box of the YOLOv5 model is directly used, the iteration speed will be very slow and the accuracy of the detector will not be high. The anchor box in YOLOv5 is obtained by the  $K$ -means clustering algorithm, which is a very classical and effective clustering algorithm. By calculating the distance (similarity) between samples, the closer samples are clustered into the same category (cluster). When

using  $K$ -means, we mainly focus on two issues: (1) the selection of cluster centers that also means dividing samples into several categories, which need to be selected according to the application scenario. (2) How to express the distance between samples? This problem also needs to be designed according to specific scenarios. Different evaluation standards have different clustering effects.

In order to detect the defects on the steel surface, first of all, we use the  $K$ -means++ algorithm to replace the original  $K$ -means algorithm in the selection of the cluster center. The  $K$ -means algorithm will randomly select the center when initializing the cluster center. If the randomly selected initial center is very poor, the iterative speed of the algorithm will be very slow, and even the wrong clustering will occur. The  $K$ -means++ algorithm will select the cluster center in the sample when initializing the cluster center, and the selected cluster centers will be as far as possible, which will eliminate the randomness of the  $K$ -means algorithm. For the selection of the anchor box for steel defect detection, when clustering algorithm is applied, the sample parameters are the length and width of the anchor frame. Therefore, the anchor box obtained by directly using the Euclidean distance as the clustering evaluation standard is not very good for steel surface defects with large differences in shape. In this paper, the IOU (Intersection over Union) of the anchor box and the bounding box is selected as the evaluation standard. A higher IOU indicates that the anchor frame is closer to the corresponding bounding box, making it more suitable for detecting the shape of steel surface defects.

**2.1.3. Model Training Steps.** The YOLOv5 model detection process mainly includes the following steps: firstly, load the relevant datasets according to the configuration file;

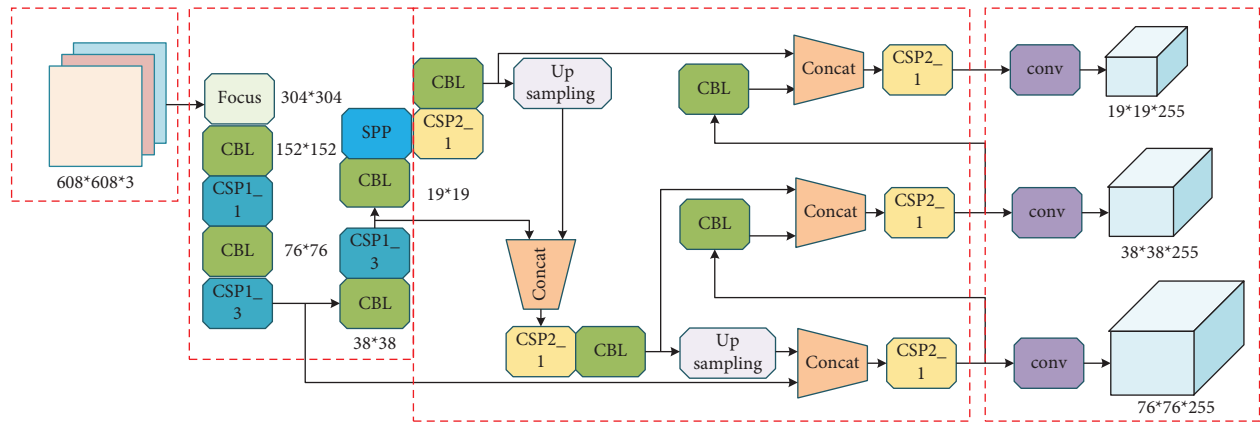


FIGURE 1: YOLOv5 network structure.

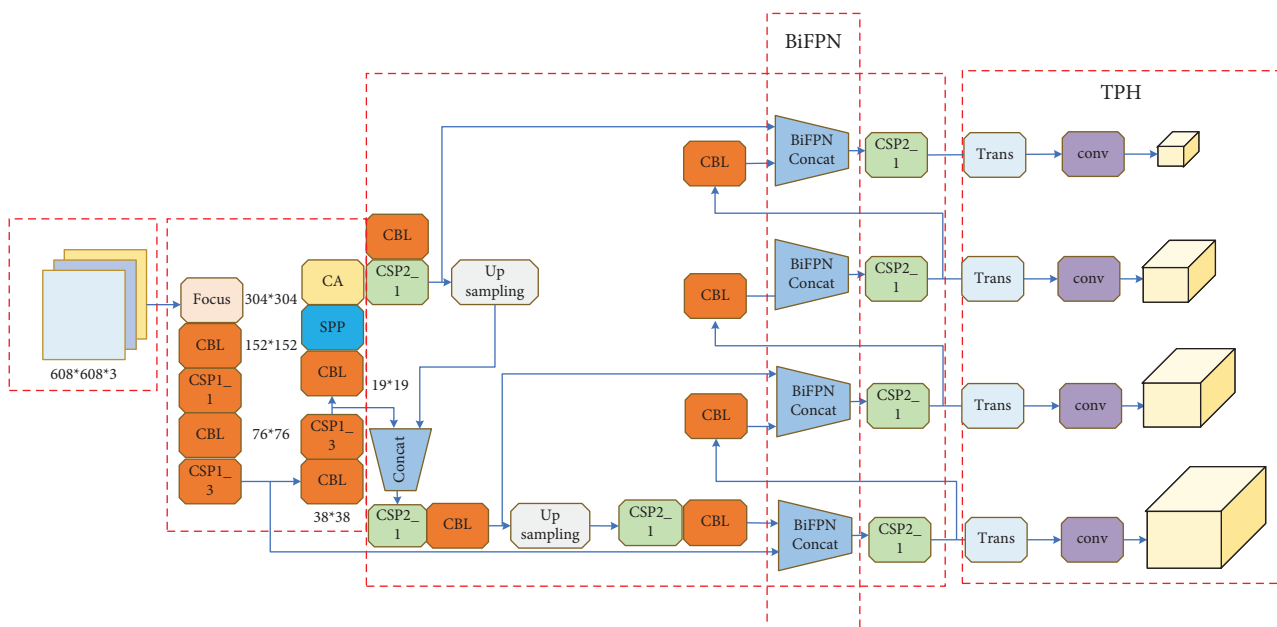


FIGURE 2: Structure of the steel surface detect model.

secondly, preprocess the dataset to meet the input requirements of the YOLOv5 model; thirdly, input the processed data into the YOLOv5 model, the model starts iterative training and updates the parameter values; finally, when the training times of the model reach the set number, the network output the final model and the training end, as shown in Figure 3.

**2.2. Coordinate Attention Mechanism.** Attention mechanism is often used to tell the model which content and which position to pay more attention to, which has been widely used in deep neural networks to enhance the performance of the model. However, in a lightweight network whose model capacity is strictly limited, most of the computing resources required by the attention mechanism are unaffordable for the lightweight network. Considering the limited computing power of the lightweight network, the SE (Squeeze-and-Excitation) [27] attention mechanism was proposed to

calculate channel attention through 2D global pooling, which provides significant performance improvement under the condition of low computing cost, but the SE attention mechanism only considers the encoding of information between channels and ignores the importance of location information. Therefore, it is very important to capture the location information of steel surface defects. The coordinate attention mechanism is specifically designed for lightweight networks. It incorporates location information into channel attention to enhance performance. The coordinate attention mechanism uses two one-dimensional global pooling operations to aggregate the vertical and horizontal input features into two independent directional awareness feature maps. Then, the two feature maps embedded with specific direction information are coded into two attention maps, each of which captures the long-range dependence of the input feature map along a spatial direction. Therefore, the location information is saved in the generated attention

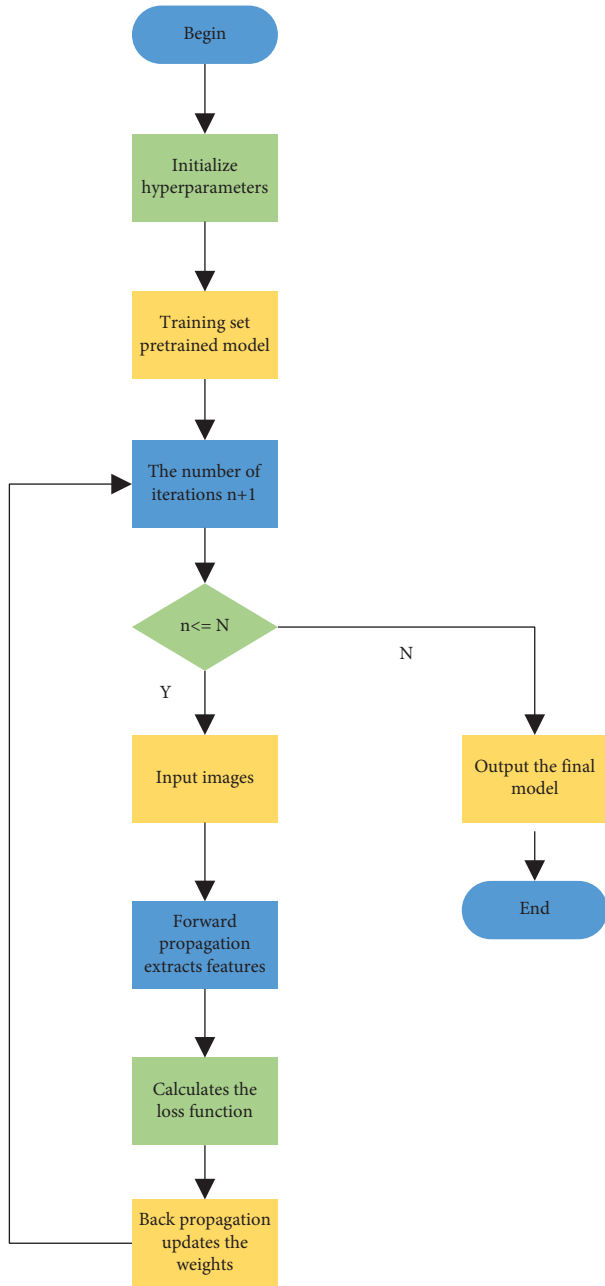


FIGURE 3: Model training process.

map, and the two attention maps are then multiplied by the input feature map to enhance the representation ability of the feature map. The network structure of the coordinate attention mechanism is shown in Figure 4.

In a highly automated steel production line, the randomness of defect causes is very small and the main causes of similar defects are determined. Therefore, for steel surface defects, especially the same type of defects, their distribution is related and the distribution position and direction of defects are similar. A coordinate attention mechanism is added to the head of YOLOv5, which makes the network model pay more attention to the spatial correlation of defects when extracting features, so as to achieve faster convergence and better detection performance of the network model.

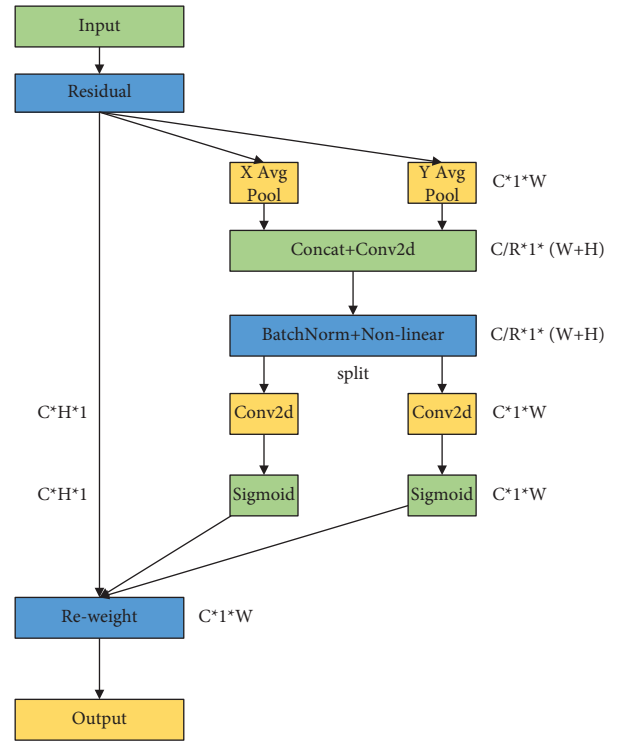


FIGURE 4: Structure of coordinate attention mechanism.

**2.3. Feature Fusion.** When detecting steel surface defects, it is easy to make mistakes in classification. The main reason for this is that steel surface defects have great differences in appearance. There are many differences between the same kinds of defects and there are also similarities between different kinds of defects. In addition, due to the change of light and material, the gray level of the flaw image will also change. All these pose a great challenge to the detection of steel surface defects. In order to deal with such problems, it is particularly important to strengthen the feature extraction of flaws.

YOLOv5's feature fusion adopts the feature pyramid network (FPN) + path aggregation network (PANet) [27, 28] scheme. The structure of the FPN and PANet is shown in Figures 5(a) and 5(b). The FPN uses the idea of image pyramid to obtain robust semantic information from top to bottom, while the PANet structure is to convey location features from bottom to top. When fusing features, most of the input features with different resolutions are added without distinction, but in fact, the contribution of these different input features to the fused output features is often unequal. In order to solve this problem, this paper uses the bidirectional weighted feature pyramid network (BiFPN) [29] for feature fusion. The structure of the BiFPN is shown in Figure 5(c). The BiFPN introduces learnable weights to learn the importance of different input features and repeatedly applies top-down and bottom-up multiscale feature fusion. Traditional feature fusion usually only overlays feature maps without distinguishing between feature maps added simultaneously. However, different feature maps have different resolutions, and their contributions to feature

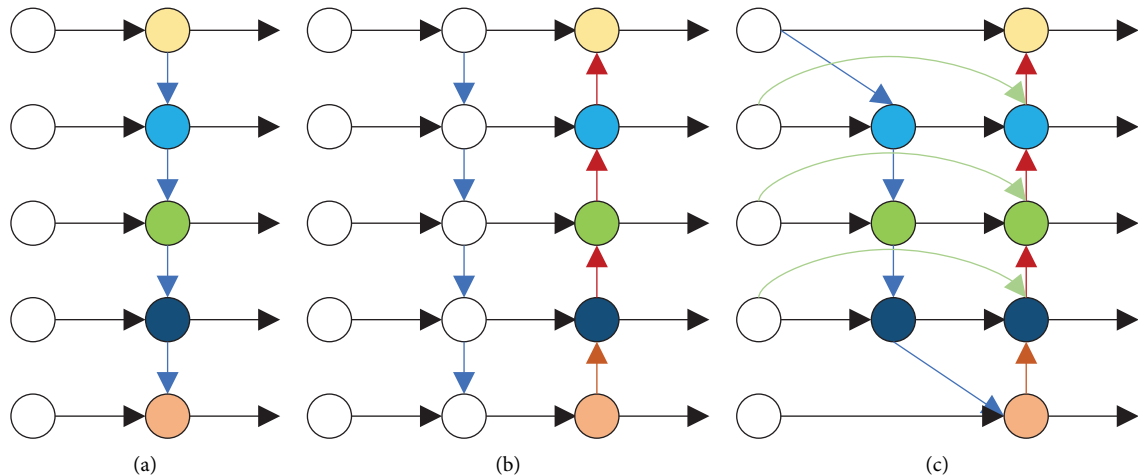


FIGURE 5: Structure of three feature fusion networks: (a) FPN, (b) PANet, and (c) BiFPN.

fusion are not equal. Therefore, simply adding feature maps is clearly not appropriate. Therefore, the BiFPN weights the features before fusion, and the fusion calculation method is shown in equation (2). Compared with the FPN, NAS-FPN, and other feature fusion networks, the PANet achieves better accuracy, but requires more parameters and calculations. As can be seen from the figure, the BiFPN removes nodes with only one input side to simplify the two-way network, and unlike the PANet, which has only one top-down and one bottom-up path, the BiFPN regards each pair of two-way paths as the same feature network layer and can repeat the same layer, which enables the BiFPN to achieve more advanced feature fusion.

$$\text{Output} = \sum_i \frac{\omega_i \times I_i}{\varepsilon + \sum_j \omega_j} \quad (2)$$

**2.4. Transformer Prediction Head.** The varying sizes of defects on steel surfaces can lead to detection errors, particularly with smaller defects. When there are gaps in the size of defects, it can decrease the accuracy of detecting small targets. YOLO series algorithms do not carry out regional sampling, so they have good performance in global information but low accuracy in small target detection. According to the characteristics of size difference, this paper adds an additional detection layer at the head of the detection network, corresponding to the detection of small resolution targets, and uses the network prediction head based on the transformer encoder [30] to improve the detection accuracy of the network model for small size defects.

The structure of the transformer encoder is shown in Figure 6. Compared with the neck module of the original YOLOv5, the transformer encoder can not only capture full text information, but also obtain rich context information. Each transformer encoder contains two sublayers. The first sublayer is the multihead concern layer and the second sublayer is the fully connected layer. There are residual connections between each sublayer. The transformer encoder can not only improves the ability to capture different

local information, but also enhances the prediction ability of the target detection network. In this paper, transformer encoder blocks are used in the head of the YOLOv5 to form the transformer prediction head (TPH) because the feature map at the end of the network have low resolution, and the application of TPH on the low resolution feature map can reduce the computing cost and improve the storage efficiency.

**2.5. Defect Size Measurement.** The steel surface defect detection system designed in this paper includes a defect size measurement method. To accurately measure the size of a defect, its edges must first be detected. Edge detection in an image is critical for object segmentation and positioning within the image. Edge detection greatly reduces the amount of data in the source image, eliminates the information irrelevant to the object, and retains the important structural attributes of the image. In order to extract more complete edges, we used the Canny detection algorithm. The Canny detection algorithm uses Sobel operator twice and uses two different thresholds to detect strong edges and weak edges, respectively. Compared with the Sobel algorithm, the Canny detection algorithm is slower, but it can identify edges with different characteristics. The result of the Canny detection algorithm is shown in Figure 7.

According to the picture, it can be found that the Canny operator can extract the defect contour more clearly, the information expression is simple, clear, and the interference is less.

Due to the reason that the shape of the defect is irregular, we take the size of the small external rectangle of the defect as the size of the defect itself. After the complete edge of the defect is extracted, the minimum external rectangle of the defect can be obtained. In this paper, a standardized reference is placed on the detect tray of the detection system. The reference is the given 10 \* 10 mm square mold with an accuracy of 0.05 mm, as shown in Figure 8. Subsequently, the actual size of each pixel in the image is determined by measuring the size of a reference object within the image.

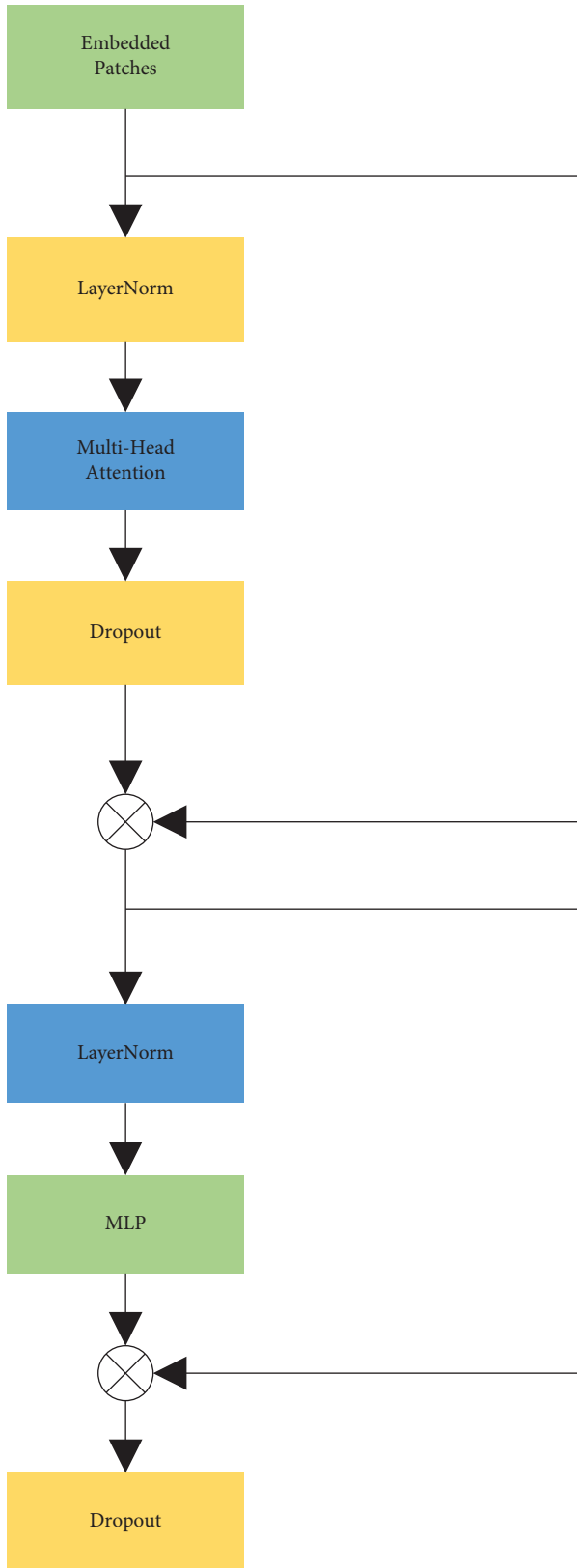


FIGURE 6: Structure of transformer encoder.

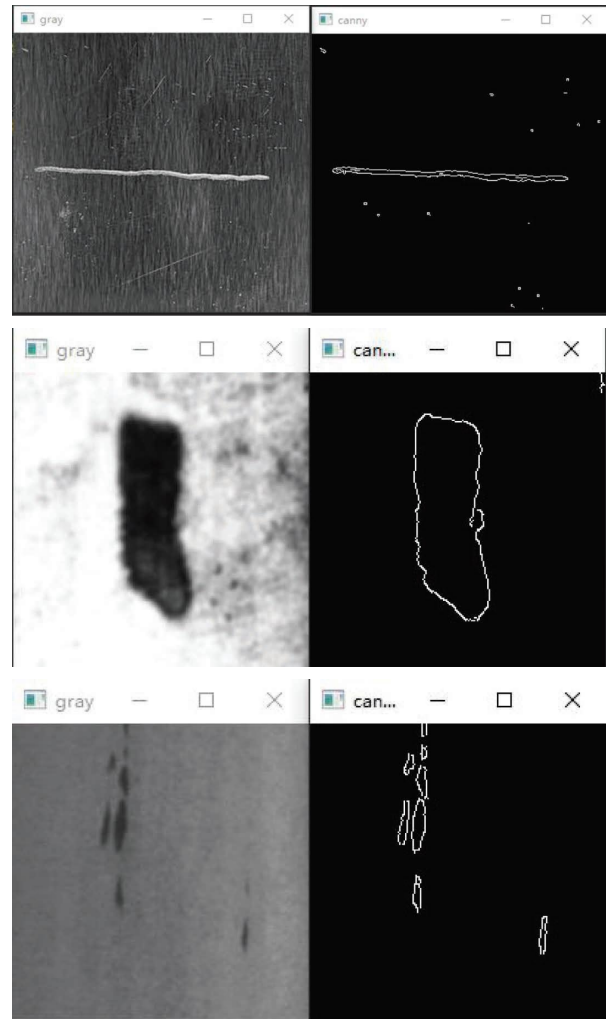


FIGURE 7: Canny edge detection result.

Finally, the number of pixels occupied by the minimum bounding rectangle of the defect is calculated, and the actual size of the detected defect is obtained through the conversion ratio between the actual size and the pixel size.

**2.6. Steel Surface Defect Detection System.** The surface defect detection system based on machine vision proposed in this paper is to detect three main defects on the steel surface and measure the defect size. The whole process is shown in Figure 9.

The hardware of the steel surface defect detection system designed in this paper is shown in Figure 10.

On the left is the acquisition module. In order to avoid the interference of external ambient light on the detection of steel surface defects, the acquisition module is carried out in a light-shielded internal environment. The acquisition module includes a CCD camera, a light source, and the steel being tested. The upper computer of the detection system

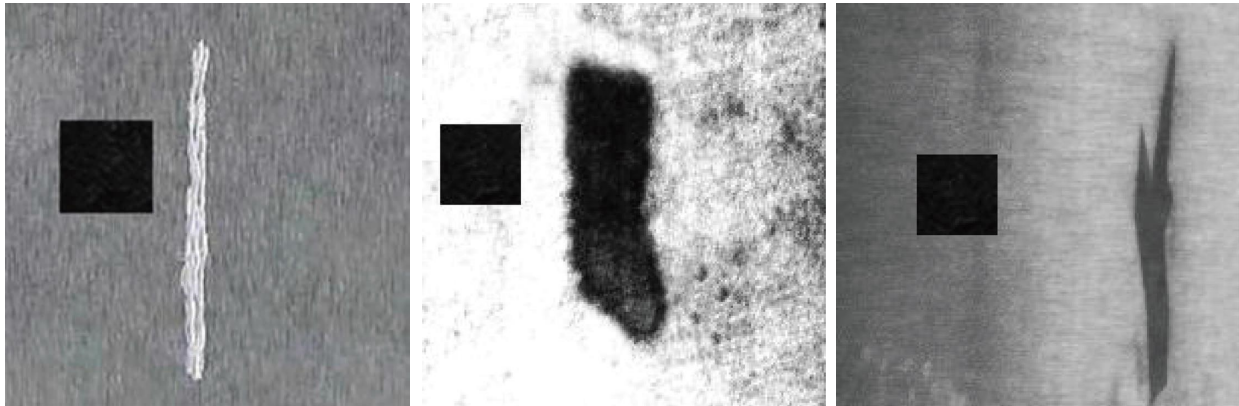


FIGURE 8: Size measurement reference and different defects.

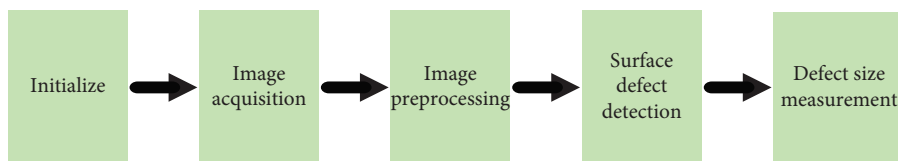


FIGURE 9: Defect detection process.

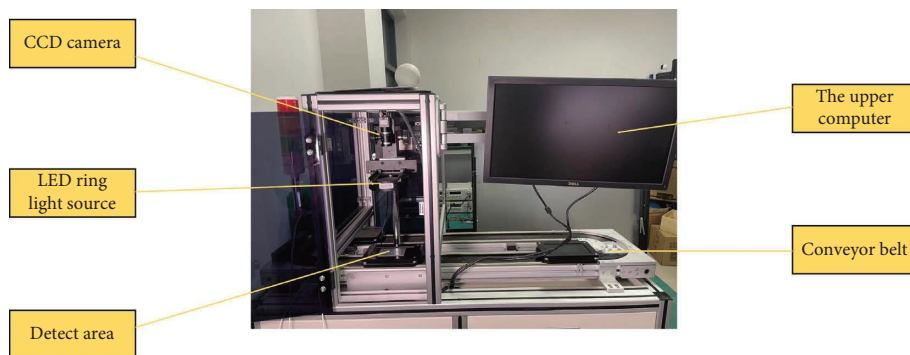


FIGURE 10: Hardware of the steel surface defect detection system.

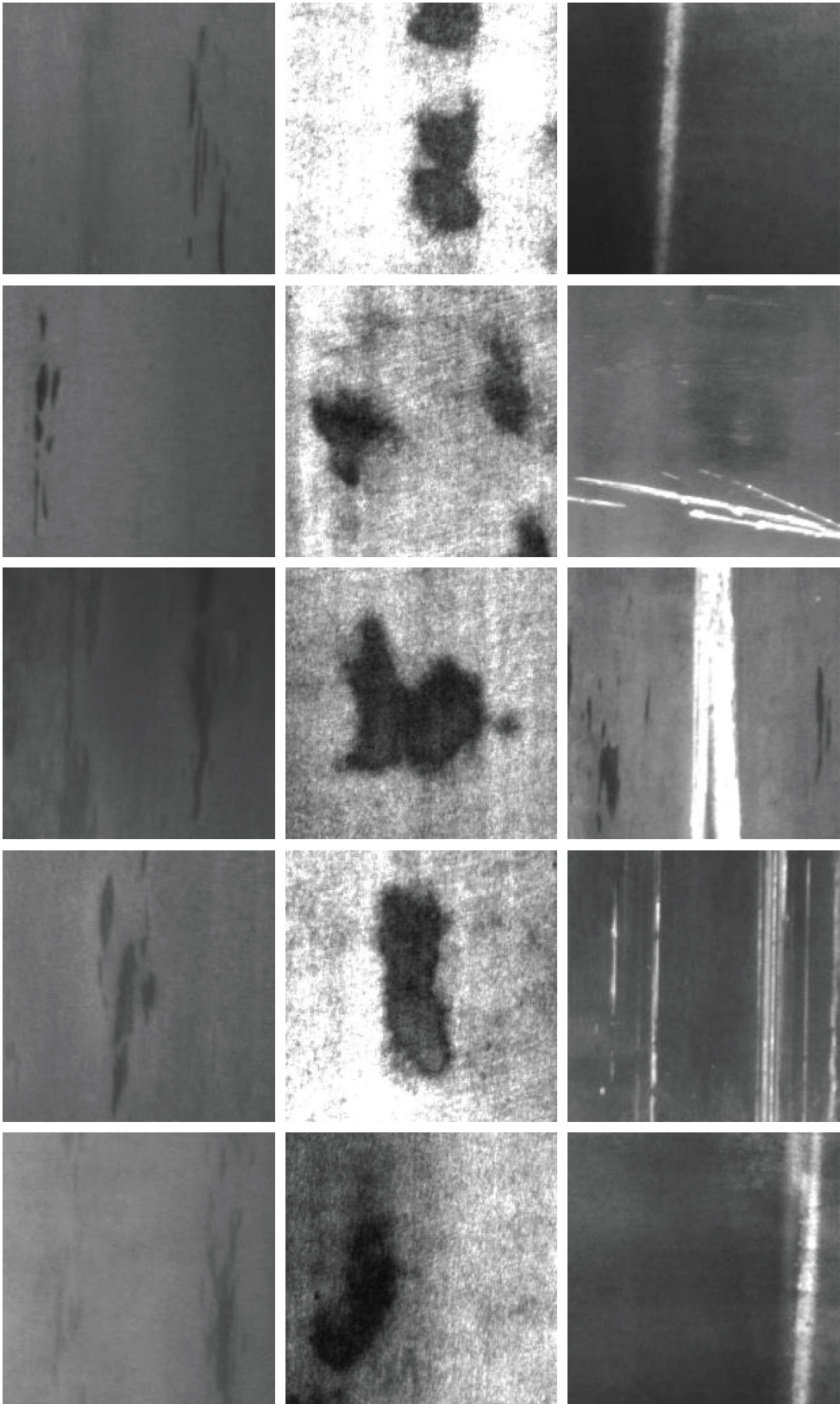
can control the detection process and view the detected defect results. The tested steel can be moved out of the detection area through the conveyor belt. The quality of the image has a significant impact on the effect of image detection, so we made a detailed analysis and comparison of the hardware cameras and light sources involved in the image acquisition module and finally determined the equipment model used in this system.

### 3. Experimental Results and Analysis

The hardware configuration used in the experiment is Ubuntu 18.04 operating system, Intel (R) Core (TM) I9-9900K processor, NVIDIA RTX3090 independent graphics card. The development framework of deep learning is Pytorch 1.13.

**3.1. Experimental Dataset.** In order to verify the practicality of defect detection and defect size measurement proposed in this paper, NEU-DET dataset is used for experiments. NEU-DET surface defect detection dataset released by the Northeast University contains gray images of surface defects of hot rolled steel strips, including three main surface defects of hot rolled steel strips, namely, patches (PA), inclusion (IN), and scratches (SC). The NEU-DET dataset is shown in Figure 11. Each defect has 300 pictures with a resolution of  $200 * 200$ . There are totally 900 images of three defects in the NEU-DET dataset. For defect detection tasks, the dataset provides labels, marking the defect category and location in each image. However, the defects in each image are not only composed of a single defect in the sample images. Also, there are similarities between different defects, which bring challenges to the inspection task.





Patches

Inclusion

Scratches

FIGURE 11: NEU-DET dataset.

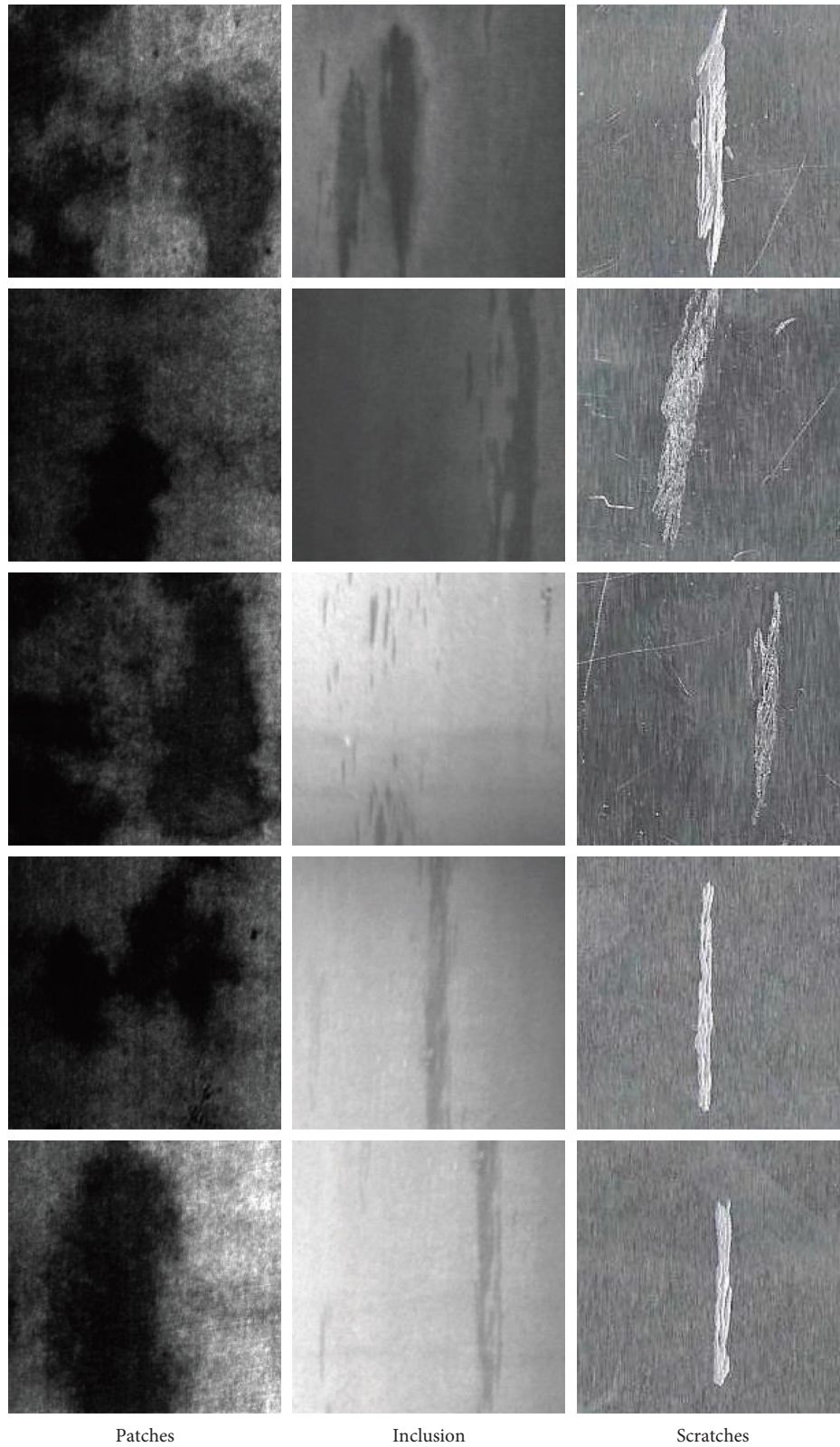


FIGURE 12: The self-made dataset.

In this paper, three types of defects in the NEU-DET dataset are selected for testing. In order to obtain better generalization ability, this paper also makes a self-made

dataset combining with the NEU-DET dataset to train the model. The steel surface defect dataset made in this paper was photographed with the selected MV-EM200 C

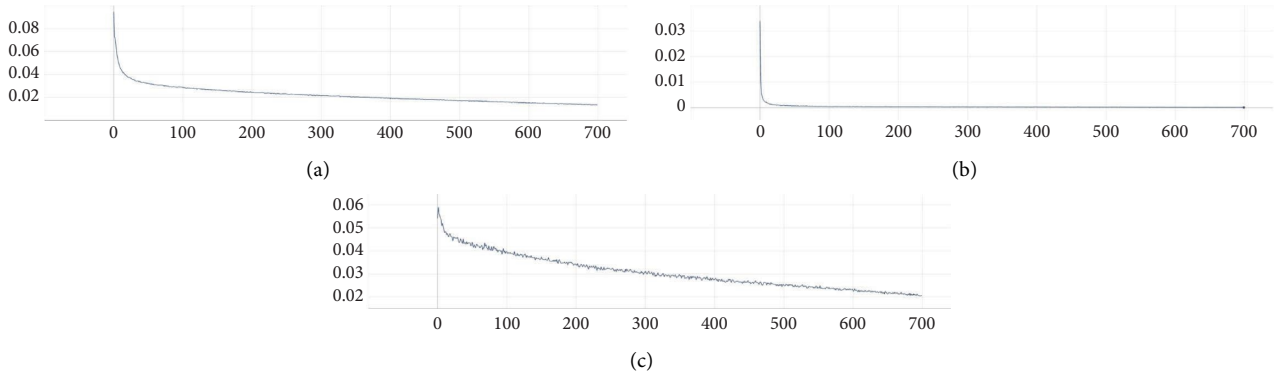


FIGURE 13: Loss function curve: (a) Box\_loss, (b) Cls\_loss, and (c) Obj\_loss.

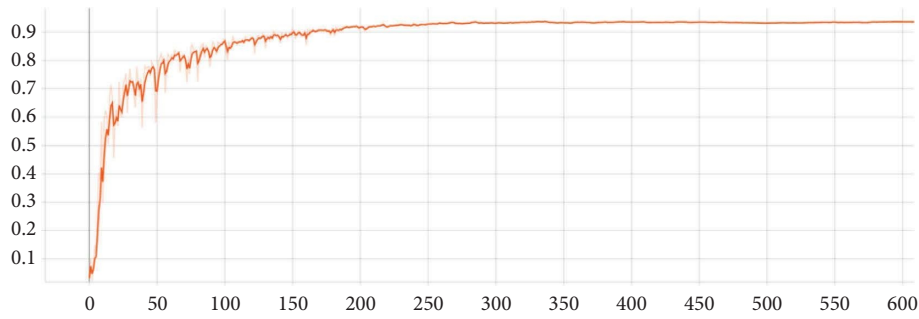


FIGURE 14: mAP\_0.5 curve.

TABLE 1: mAP\_0.5 data comparison.

	mAP_0.5 (ALL) (%)	mAP_0.5 (IN) (%)	mAP_0.5 (PA) (%)	mAP_0.5 (CR) (%)
YOLOv5	90.4	90.2	86.9	94
YOLOv5 + TPH	90.8	90.6	87.1	93.7
YOLOv5 + BiFPN	91.2	89.8	88.4	94.1
Ours	93.6	93.3	91.8	95.7

TABLE 2: mAP\_0.5 comparison of attention mechanisms.

	mAP_0.5 (ALL) (%)	mAP_0.5 (IN) (%)	mAP_0.5 (PA) (%)	mAP_0.5 (CR) (%)
YOLOv5 + CBAM	89.3	88.4	85.9	93.6
YOLOv5 + SE	90.5	89.6	88.1	93.9
YOLOv5 + CA (ours)	93.6	93.3	91.8	95.7

camera. We take pictures by adjusting the size and clarity of the window, the focal length of the lens, the angle, and brightness of the subject to an appropriate position. Through the abovementioned steps, the image of the workpiece can be collected by the CCD camera and LED ring light source. For 210 pictures collected in this paper, in order to solve the problem of overfitting caused by small amount of picture data, we conduct data augmentation processing on the images. To increase the quantity of available data, we applied image flipping, random clipping, and RGB channel exchange methods. This process expanded the amount of defect picture data

to ten times the original size for our experiment. The experiment proved that it played a good role in the follow-up model training and result testing, especially the generalization ability of the model was improved. Through the abovementioned data augmentation technology, we expanded the dataset to 2100 samples. More samples are conducive to improving the generalization ability of the model. We made the dataset on the basis of 2100 pictures. The self-made dataset used for training is shown in Figure 12. The original image has no location coordinate of the area where the defect is located, which needs to be manually marked. Use Labelmg software to

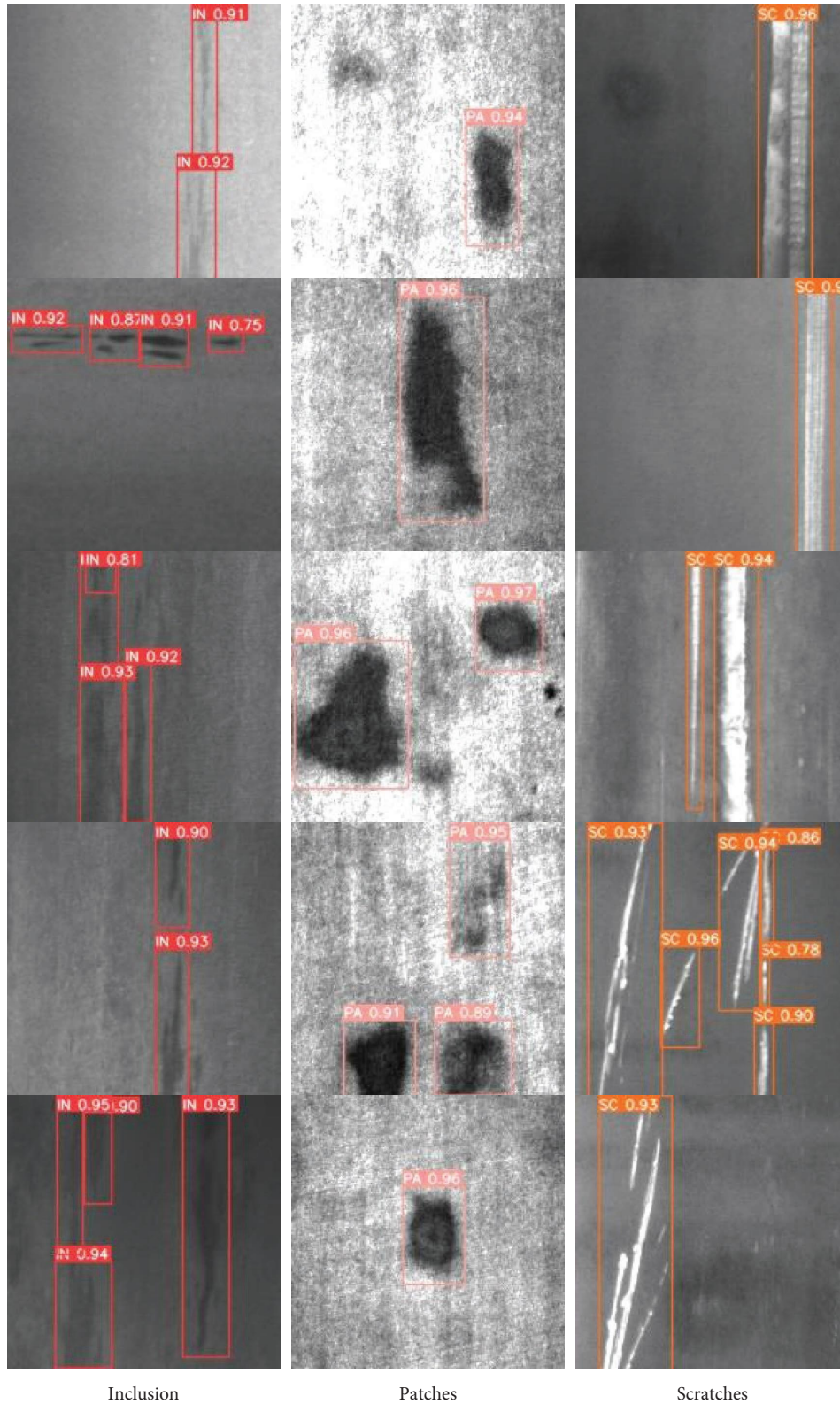


FIGURE 15: Defect detection.

mark the defects on the picture, that is, mark the location coordinates of the defects and the types of defects. After marking, divide all 3000 defect pictures in the self-made

dataset and NEU-DET dataset into training set and test set in a 4:1 ratio. There are 2400 pictures in the training set and 600 pictures in the test set.

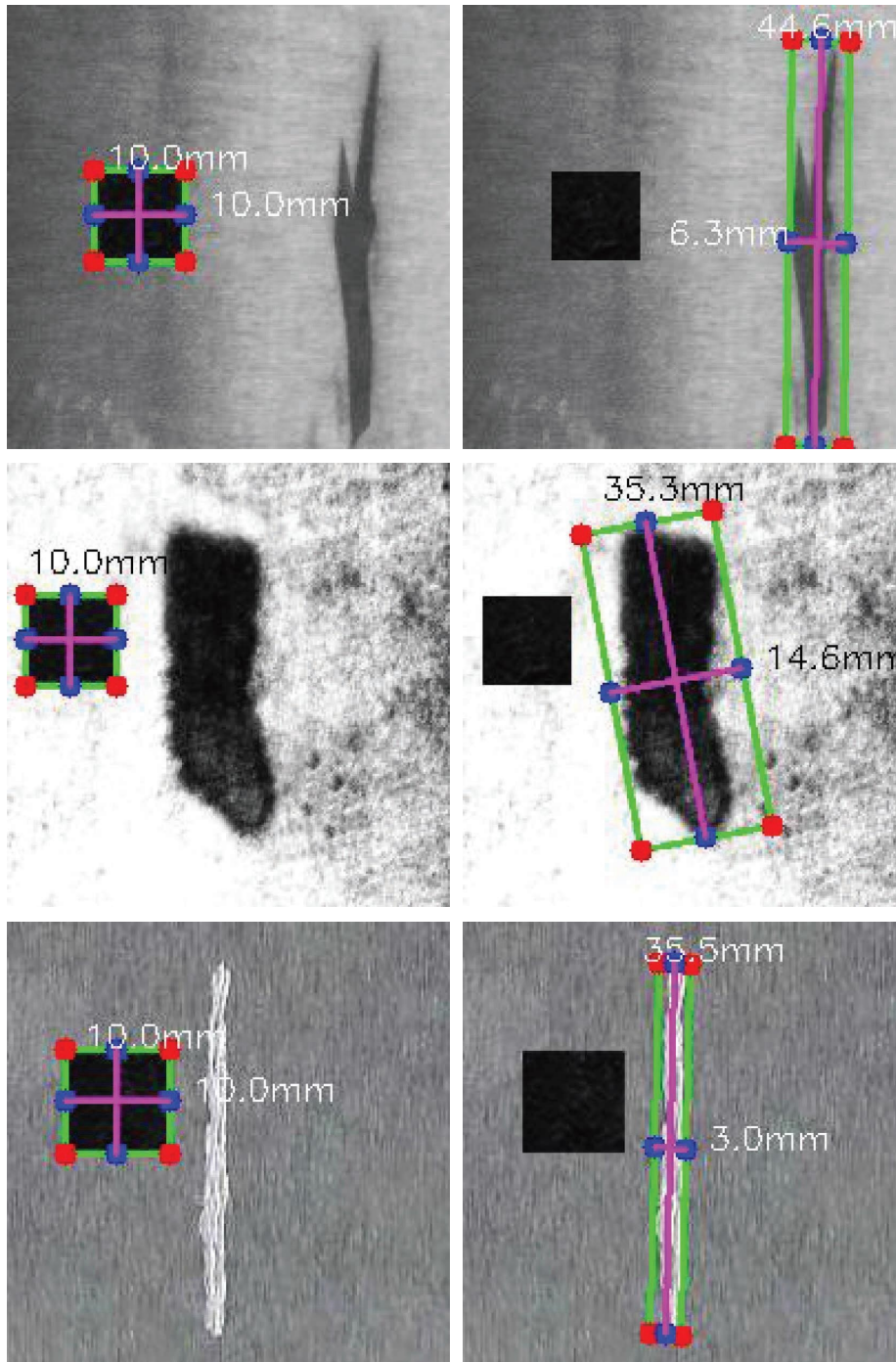


FIGURE 16: Defect size detection.

3.2. Model Training and Analysis. We build an improved YOLOv5 model through the deep learning framework. First, we conduct pretraining on the COCO dataset and then use the pretraining results to train the defect image dataset. The image size of the model input is  $416 * 416$ . During training, the batch size was set to 32, and 500 epochs were learned in total. The training loss function is shown in Figure 13. In the

figure, the upper is the loss of the bounding box, the middle is the classification loss, and the lower is the confidence loss.

The figure illustrates that the value of the loss function gradually decreases and approaches convergence with an increasing number of iterations. Notably, the classification loss reaches convergence the fastest, followed by boundary box loss, while confidence loss convergence is the slowest.

This may be due to the large difference between the same kinds of defects.

AP (average accuracy) is the average of accuracy rates under different recall rates, which is usually used to evaluate the accuracy of the model. mAP (mean average accuracy) is the average detection accuracy of all target categories. It is usually used to evaluate the overall performance of the model. It is a commonly used comprehensive indicator in the field of object detection. It measures the overall detection accuracy of the detection frame under different IOUs. The larger the value, the higher the model accuracy. mAP is calculated using precision and recall. The calculation formula of mAP is shown in equations (3)–(6):

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (3)$$

$$\text{recall} = \frac{TP}{TP + FN}, \quad (4)$$

$$AP = \int_0^1 \text{precision}(\text{recall})d(\text{recall}), \quad (5)$$

$$\text{mAP} = \frac{\sum_{i=0}^n AP(i)}{n}. \quad (6)$$

Among these, true positive (TP) examples are positive examples that are correctly predicted, false positive (FP) examples are negative examples that are wrongly predicted as positive examples, and false negative (FN) examples are positive examples that are wrongly predicted to be negative examples, where  $n$  is the number of detection categories and AP is the detection accuracy of various types, where AP can be expressed as the area of the curve made with recall as the horizontal axis and precision as the vertical axis, that is, the area of the PR curve is calculated using the integral formula.

As can be seen from Figure 14, the mAP<sub>0.5</sub> value increased with the increase of training times and the average accuracy achieved 93.6%. In order to verify the performance of this paper, this paper also conducted a comparative experiment with YOLOv5, and the results are shown in Table 1. Compared with the original YOLOv5 model, the overall accuracy of this paper is improved by 3%, among which the accuracy of scratches is improved by 5%, with the most obvious effect.

### 3.3. Comparison of Average Precision with Different Methods.

This section shows the comparison among three kinds of attention mechanism. They are Squeeze-and-Excitation (SE) network, convolutional block attention module (CBAM), and coordinate attention (CA). According to Table 2, it is obvious that the steel surface defect detection method proposed in this paper is better than the SE and the CBAM. The figure shows that the overall mAP and each mAP of the defects of the proposed model are better than the others by a 5% difference.

**3.4. Defect Detection.** In this paper, the trained YOLOv5 model is deployed to the experimental platform. The detection results of inclusions, scratches, and patches are shown in Figure 15.

The figure reveals that the model has successfully detected all three types of defects and labeled their respective locations with rectangular boxes, indicating promising results. These outcomes demonstrate that the model can accurately distinguish and identify various defects, even if they appear simultaneously on the same steel surface. The detection accuracy of three kinds of defects is more than 90%, and the detection accuracy of individual defects may be low. Through a large number of experiments, we found that the low accuracy of individual defects may be attributed to the undersizing of the defects.

It can be seen from the figure that we can calculate the actual size of the defect through the external rectangle and scale conversion of the defect.

**3.5. Defect Size Measurement.** Because the NEU-DET dataset does not provide data of steel, camera, and other experimental conditions, the defect size measurement experiments proposed in this paper are all conducted in the self-made dataset. In order to extract complete defects, the Canny edge detection algorithm is first used to obtain the boundary of defects, then the minimum circumscribed rectangle function box in openCV image processing library is called to select the location of defects, and then the actual size of defects is calculated by using the conversion between the actual size and pixel size to determine the proportion. The detection effect is shown in Figure 16.

It can be seen from the figure that we can calculate the actual size of the defect through the external rectangle and scale conversion of the defect.

## 4. Conclusion

This paper designs and deploys a surface defect detection and defect size measurement system based on the YOLOv5 model. On the one hand, in order to prove the performance of the steel surface defect detection algorithm proposed in this paper, we have verified it through the NEU-DET dataset and self-made dataset and obtained the following conclusions: (1) a method of feature fusion is applied, instead of the original FPN + PANet frame. Due to the large difference in shape of steel surface defects, this paper uses the BiFPN to obtain better performance. (2) The coordinate attention mechanism is added to the head of YOLOv5. The detection speed is not affected while the model parameters are increased. The accuracy is improved by about 3% compared with the original YOLOv5. (3) A new prediction head consist of the transformer encoder is also proposed in this paper; it can improve the accuracy of the classification. On the other hand, in order to ensure that high-quality pictures can be collected in the actual production environment, this paper

compares and analyzes a variety of major image acquisition devices in detail, comprehensively considers the cost and performance factors of the devices, and selects relevant devices scientifically. In order to make the defect detection model have better generalization performance, this paper makes the model reach 93.6% mAP and reasoning speed reach 133FPS by means of data augmentation, migration learning, and other methods. The system designed and deployed in this paper can achieve the effect of real-time detection while ensuring high accuracy and provides a feasible scheme for eliminating the products with surface defects on the pipeline.

In the process of improving the model, there are also some problems. Industrial testing involves a wide range of detection scenarios, and as such, we aim to develop a detection model that is less dependent on image resolution. For instance, motion blur may occur when capturing images of objects in motion on a conveyor belt. To address this issue, we plan to leverage generative adversarial networks (GANs) to enhance and refine the blurred images. In addition, to improve the detection of small defects, we may need to increase the number of detection layers, but this could potentially increase the parameter quantity of the network model. Thus, deciding on the optimal trade-off between detection accuracy and data volume will be one of the key challenges that we aim to tackle in our future work.

### Data Availability

The data that support the findings of this study are available from the author (Ziheng Ding) upon reasonable request.

### Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

### Acknowledgments

This research was funded by the National Natural Science Foundation of China (Grant No.61973178 and 62103205), the smart grid joint fund of state key program of the National Natural Science Foundation of China (Grant No. U2066203), the Major Natural Science Projects of Colleges and Universities in Jiangsu Province (Grant No. 21KJA470006), and the Postgraduate Research & Practice Innovation Program of Jiangsu Province (Grant No. KYCX22\_3347).

### References

- [1] F. Deng, X. Zhang, Y. Ning, and L. Zhao, "An Improved RAPID Imaging Method of Defects in Composite Plate Based on Feature Identification by Machine Learning," *Sensors*, vol. 22, no. 21, p. 8413, 2022.
- [2] F. Wu, Q. Li, S. Li, and T. Wu, "Train Rail Defect Classification Detection and its Parameters Learning Method," *Measurement*, vol. 151, Article ID 107246, 2020.
- [3] Y. Liu, K. Xu, and J. Xu, "An Improved MB-LBP Defect Recognition Approach for the Surface of Steel Plates," *Applied Sciences*, vol. 9, no. 20, p. 4222, 2019.
- [4] T. Zimmermann, G. Ciuti, M. Milazzo et al., "Visual-based defect detection and classification approaches for industrial applications—a survey," *Sensors*, vol. 20, no. 5, p. 1459, 2020.
- [5] J. Sun, C. Li, X. Wu, V. Palade, and W. Fang, "An effective method of weld defect detection and classification based on machine vision," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 12, pp. 6322–6333, 2019.
- [6] S. E. Park, S.-H. Eem, and H. Jeon, "Concrete crack detection and quantification using deep learning and structured light," *Construction and Building Materials*, vol. 252, Article ID 119096, 2020.
- [7] W. Zhao, F. Chen, H. Huang, D. Li, and W. Cheng, "A New Steel Defect Detection Algorithm Based on Deep Learning," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 5592878, 13 pages, 2021.
- [8] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, 2015.
- [9] Y. Wu, D. Guo, H. Liu, and Y. Huang, "An End-To-End Learning Method for Industrial Defect Detection," *Assembly Automation*, vol. 40, no. 1, pp. 31–39, 2019.
- [10] J. Guan, J. Li, X. Yang, X. Chen, and J. Xi, "Defect Detection Method for Specular Surfaces Based on Deflectometry and Deep Learning," *Optical Engineering*, vol. 61, no. 6, Article ID 61407, 2022.
- [11] X. Kou, S. Liu, K. Cheng, and Y. Qian, "Development of a YOLO-V3-based Model for Detecting Defects on Steel Strip Surface," *Measurement*, vol. 182, Article ID 109454, 2021.
- [12] D. He, K. Xu, and P. Zhou, "Defect Detection of Hot Rolled Steels with a New Object Detection Framework Called Classification Priority Network," *Computers & Industrial Engineering*, vol. 128, pp. 290–297, 2019.
- [13] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2014, <https://arxiv.org/abs/1409.1556>.
- [14] K. Ihor, M. Pavlo, B. Vitaly, and P. Olegas, "Recognition of Scratches and Abrasions on Metal Surfaces Using a Classifier Based on a Convolutional Neural Network," *Metals*, vol. 11, no. 4, p. 549, 2021.
- [15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, June 2016.
- [16] M. Sharma, J. Lim, and H. Lee, "The Amalgamation of the Object Detection and Semantic Segmentation for Steel Surface Defect Detection," *Applied Sciences*, vol. 12, no. 12, p. 6004, 2022.
- [17] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Berlin, Germany, November 2015.
- [18] Y. Xie, W. Hu, S. Xie, and H. Lei, "Surface Defect Detection Algorithm Based on Feature-Enhanced YOLO," *Cognitive Computation*, vol. 15, pp. 565–579, 2022.
- [19] Y. Zhang, X. Liu, J. Guo, and P. Zhou, "Surface Defect Detection of Strip-Steel Based on an Improved PP-YOLOE-m Detection Network," *Electronics*, vol. 11, no. 16, p. 2603, 2022.
- [20] S. Guan, M. Lei, and H. Lu, "A Steel Surface Defect Recognition Algorithm Based on Improved Deep Learning Network Model Using Feature Visualization and Quality Evaluation," *IEEE Access*, vol. 8, 2020.

- [21] C.-C. Yeung and K.-M. Lam, "Efficient Fused-Attention Model for Steel Surface Defect Detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, 2022.
- [22] Y. Liu, K. Xu, and J. Xu, "Periodic Surface Defect Detection in Steel Plates Based on Deep Learning," *Applied Sciences*, vol. 9, no. 15, p. 3127, 2019.
- [23] X. Shi, Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. Woo, "Convolutional LSTM network: a machine learning approach for precipitation nowcasting," *Advances in Neural Information Processing Systems*, vol. 28, 2015.
- [24] P. Damacharla, M. V. Achuth Rao, R. Jordan, and A. Y. Javid, "TLU-net: a deep learning approach for automatic steel surface defect detection," in *Proceedings of the 2021 International Conference on Applied Artificial Intelligence*, Halden, Norway, May 2021.
- [25] L. Yang, X. Wei, Y. Li, Z. Dong, and M. Shahidehpour, "Detection of False Data Injection Attacks in Smart Grid: A Secure Federated Deep Learning Approach," *IEEE Transactions on Smart Grid*, pp. 4862–4872, 2022.
- [26] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, TN, USA, June 2021.
- [27] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, June 2018.
- [28] S. Liu, Q. Lu, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, June 2018.
- [29] M. Tan, R. Pang, and V. Quoc, "Efficientdet: scalable and efficient object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, June 2020.
- [30] X. Zhu, S. Lyu, X. Wang, and Z. Qi, "TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Montreal, Canada, October 2021.