

Research Article

A Novel Defect Detection Method for Ferrite Shield Surface Defects by Improved Faster R-CNN

Xuhui Ye ^{1,2}, Jun Ye ¹, Zhuang He ¹, Daode Zhang ¹, Xinyu Hu ¹ and Qi Chen ¹

¹School of Mechanical Engineering, Hubei University of Technology, Wuhan 430068, China

²Collaborative Innovation Center of Intelligent Green Manufacturing Technology and Equipment, Qingdao 266000, China

Correspondence should be addressed to Daode Zhang; 19951044@hbut.edu.cn

Received 4 March 2022; Revised 15 October 2022; Accepted 29 October 2022; Published 9 November 2022

Academic Editor: Sadiq Hussain

Copyright © 2022 Xuhui Ye 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.

Traditional manual inspection of small objects such as ferrite shield surface defects through naked eyes is labour-intensive and inefficient, especially existed false and missed inspection. To improve the defect detection rate and efficiency, a novel inspection method based on an improved Faster R-CNN is proposed aiming at weak feature information, small object, and diverse shapes of the defective target. This paper takes cracks, pits, impurities, and dirty defects existing on the surface of the ferrite shield as examples to verify this method. Firstly, the bidirectional feature fusion network with ResNet-50 is added to the original Faster R-CNN to obtain feature maps, which contain strong semantic information and rich location information of the defects. Then, *k*-means clustering with genetic algorithm is adopted for generating adaptive anchor boxes to match defects of various shapes in the region proposal network (RPN) stage. Then, the region of interest (ROI) Align instead of the ROI Pooling is introduced to eliminate the candidate frame position bias and improve the defect localization accuracy. Finally, Soft nonmaximum suppression (NMS) is used to reduce the probability of defective targets. The mean average precision (mAP) of the applied method for detecting surface defects of ferrite shields is 81.0% which is higher than the commonly used detectors such as SSD, YOLO, RetinaNet, and Cascade RCNN. The study can provide a new detection mean to obtain the detection capability and accuracy of ferrite shields and lay a solid foundation for automated detection of small objects.

1. Introduction

Ferrite shield, also known as magnetic ceramics, is commonly used in the field of radar, communication, and electronic computers due to its good magnetic permeability. It can shield the external electromagnetic waves from the internal circuits. Its shielding performance is greatly affected by the surface quality, so it is necessary to accurately and efficiently detect defects on the surface of the ferrite shield. Since the current manual detection with the aid of optical equipment is inefficient and unreliable, it was gradually replaced by the machine vision solution, which is faster, more efficient, more practical, and less costly.

Ferrite shield surface defects mainly contain crazing, pit, impurity, and dirt, which are characterized by weak feature information, small object, and diverse shapes. The current

methods for detecting defects on the surface of industrial products can mainly be divided into traditional image processing methods and deep learning methods.

Traditional methods generally separate the defects from the background by threshold segmentation first, and the type of defects is determined by the pattern recognition algorithm. Fan et al. proposed a visual detection method for defective fasteners based on the region features [1]. The region of interest extraction and template matching algorithm is used to detect defects of railroad fasteners in complex backgrounds. Wang and Zhang proposed a defect detection algorithm for X-ray images of automotive wheel castings [2]. The subtle defects in wheel castings are identified by using the pattern recognition technology. Yuan et al. proposed a rail defect segmentation algorithm based on the OTSU method [3], using the interclass variance threshold segmentation method with target variance

weighting to complete the segmentation detection of small and medium defect targets in the large background region of rails. Yang and Dai proposed a local image alignment method for food packaging printing defect detection [4], combining wavelet transform and edge detection algorithm to detect microdefects, such as knife fuse and braces. However, these methods cannot effectively distinguish the target defect from the background due to the complex irregular texture of ferrite shielding surface and the low contrast between the small target defect and the background. Moreover, the performance of traditional methods is easily affected by the ambient lighting condition and time-consuming, which cannot provide stable and fast detection performance.

The detection methods based on deep learning can simplify the image processing process by automatically extracting image features. Due to its powerful adaptability and efficiency of visual detection, they have become the main methods for defect detection in the industry.

Mei et al. proposed a multiscale convolutional denoising autoencoder network model [5], which uses encoding and decoding to reconstruct defects, enabling the identification of fabric surface defects with a small number of defect-free samples. Tao et al. designed a cascaded autoencoder structure that introduced a semantic segmentation-based prediction mask and predicted the classification of segmentation results by convolution [6]. Yang et al. proposed a multiscale SSD network image detection algorithm to improve the detection ability of tiny target defects with low-resolution images [7]. Fadli and Herlistiono used a CNN model with the Xception architecture to accomplish high frame rate defect detection for high-resolution steel images [8]. Wang et al. built a deep CNN network model and obtained a high detection accuracy [9]. He et al. added a multilevel feature fusion network (MFN) on the basis of the CNN model [10] and combined with the region proposals network to achieve 82.3 mAP on the NEU-DET dataset.

As a solution, the Faster R-CNN achieves object detection performance with high accuracy through a two-stage network with the RPN, which has displayed competitive performance in defect detection. At present, the Faster R-CNN has been extensively used in industry detection fields. Ren et al. proposed a lightweight network based on Faster R-CNN using separable convolution and adding central loss to the original loss function to achieve 98.32 mAP on strip surface defect detection [11]. Zhao et al. proposed a reconstructing Faster R-CNN network with multiscale fusion for the problem of defect detection in steel production and the mAP is 0.752 [12]. Hu and Zeng proposed a Faster R-CNN algorithm with the feature pyramid network (FPN) combined with region proposal by guided anchoring (GARPN) and the atrous spatial pyramid pooling-balanced-feature pyramid network (ABFPN), respectively, to detect surface tiny defects of the printed circuit board (PCB) [13, 14]. A symmetric structured feature extraction network was proposed by Xing and Jia, which achieved 0.784 mAP on a detection dataset of surface defects on hearth of raw aluminum casting [15]. Gao et al. proposed a small sample gear face defect detection method based on a

deep convolutional generative adversarial network (DCGAN) and a lightweight convolutional neural network (CNN), which is helpful for few-shot object detection [16]. For the one-stage network, SO-YOLO used feature fusion to improve the performance of small target detection, and YOLO-C add several attention mechanism modules to enhance the detection accuracy [17, 18].

These deep learning methods can provide a reference for ferrite shield surface defect detection. However, the main research objects of the abovementioned methods are regular textured cloth and smooth steel surfaces, which are characterized by larger size and higher contrast to the target defects. Moreover, the effect of these methods for the detection of defects in this paper is open to question.

Given the above, an improved Faster R-CNN network is proposed for the detection of surface quality defects of ferrite shields. Firstly, ResNet-50 [19], characterized by a deep network structure, is adopted as the backbone of the object detector to extract features. The location information and semantic information are enhanced by a bidirectional feature fusion module [20], which can obtain a more reliable feature representation of the defective target. Secondly, in the RPN stage, the anchor boxes calculated by k -means clustering [21] with genetic algorithm [22] are used to obtain adaptive proposals, which is vital to achieve the prediction of morphologically diverse defect targets. Thirdly, ROI Align [23] is used instead of the original ROI Pooling to eliminate the bias caused by pooling rounding, and it also improves the accuracy of the system defect detection. Finally, Soft nonmaximum suppression [24] is adopted to reduce the confidence level of proposals larger than the threshold value, thus reducing the missing detection rate.

The remainder of this article is organized as follows: the construction and data distribution of the FSSD dataset are elaborated in Section 2. The structure of the improved Faster R-CNN is performed in Section 3 with an in-depth analysis of adopted strategies. In Section 4, the proposed network is applied to the ferrite shield surface defect detection task. Finally, conclusions and an outlook of future works are presented in Section 5.

2. Experimental Data

2.1. Data Collection. To verify the effectiveness of the proposed algorithm, we constructed a self-built FSSD dataset. A CCD camera of Hikvision MV-CA050-10GM with MVL-MY-05-110-MP telocentric lens is used to capture images of all shield workpiece in the laboratory. A low-angle blue light source is placed 30 mm away from the shield workpiece vertically with maximum brightness. The camera is placed about 110 mm away from the object, and the camera exposure time is set to 20 ms. The experimented equipment is shown in Figure 1.

2.2. Data Preprocessing. As shown in Figure 2, the ferrite shield occupies about 70% of the overall image due to the presence of the background. Directly feeding the whole image into the network for training will increase the

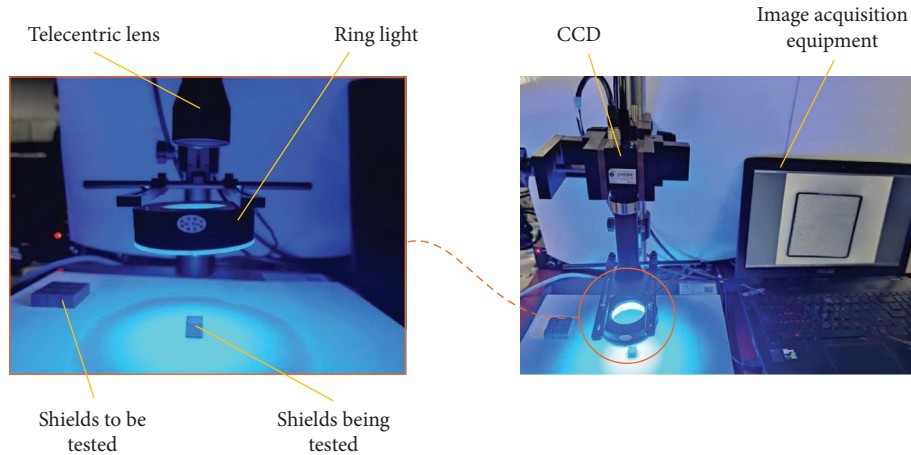


FIGURE 1: Image acquisition process.

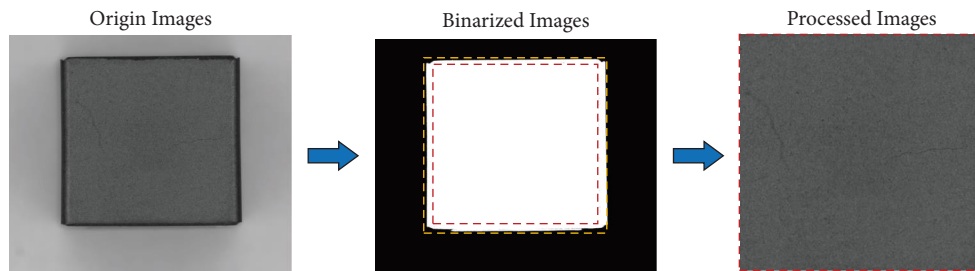


FIGURE 2: Image processing. The binarized image is obtained by using OTSU on the original image. The yellow dashed box is the outline of the workpiece and the red dashed box is the cropped useful image.

computational cost, and it may decrease the accuracy of the network. Therefore, it is necessary to preprocess the image before feeding them into the network. The overall contour image of the shield workpiece surface is extracted based on OTSU [25] and the mapping transformation algorithm [26]. Then, the image is cropped to 1200×1200 pixels, as shown in Figure 2.

One thousand high-quality images were selected from the cropped image dataset manually. As shown in Figure 3, there are four types of defects namely crazing, pit, impurity, and dirt, which are characterized by the following issues:

- (1) The texture of the ferrite surface is complex, the contrast between the defects and the background is low, and the feature information of defective targets is weak
- (2) For pit and impurity defects, they are both relatively small in the image
- (3) Defects show diverse shapes, with some being narrow and some being large, such as crazing and dirt

The image annotation tool called LabelImg is used to label the defective image according to the PascalVOC2007 dataset format. The number of the four types of defects are 928, 512, 385, and 57, respectively. The imbalance of data between categories will cause the model to be overfitted in the training process, which will directly lead to the prediction results being more inclined to crack defects. Therefore, the dataset is filtered again to exclude images

containing only single crack defects, and the images containing dirty defects are expanded by applying gamma transform. As shown in Table 1, a total of 1000 image databases containing various types of defects were acquired.

The dataset is randomly shuffled and divided according to the ratio of 7:1:2, and 700 training set images, 100 validation set images, and 200 test set images are obtained. In order to improve the generalization ability of the model, the image rotation and inversion operations are applied to increase the number of training images. Specifically, rotating the original image by 90° , 180° , and 270° and flipping in the horizontal and vertical directions led to the expansion of the number of training sets to 4200. The final number and distribution of the obtained datasets are shown in Table 2.

3. Methodology

Object detection is a technique to identify and localize object of interest in images, and deep learning-based object detection techniques are applied in various fields [27]. At present, the object detection methods based on deep learning are mainly divided into one-stage and two-stage methods [28]. One-stage detection algorithm can directly generate the category probability and position coordinate value of objects, and the representative one-stage target detection algorithms include SSD [29] and YOLO [30]. The two-stage methods divide the detection problem into two stages. In the first stage, candidate regions containing

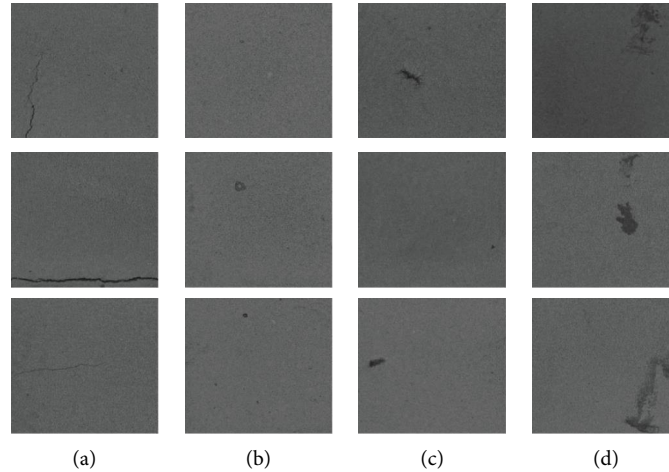


FIGURE 3: Examples of different defect types in the FSSD dataset: (a) crazing; (b) pit; (c) impurity; (d) dirt.

TABLE 1: The number of defects of the FSSD dataset.

Defects	Crazing	Pit	Impurity	Dirt
Original	928	512	385	85
Processed	561	552	438	378

TABLE 2: The amount and distribution of the FSSD dataset.

Datasets	Training sets	Validation sets	Test sets	Total number
Original	700	100	200	1000
Processed	4200	100	200	4500

approximate target location information are generated, and in the second stage, candidate regions are classified and refined. The representative two stage object detection algorithms include R-CNN [31], Fast R-CNN [32], and Faster R-CNN [33].

Faster R-CNN is adopted as the basic network in our study for its advantages over one-stage methods in accuracy. We optimize some modules on the Faster R-CNN to solve the problems of detecting defects with weak feature information, small object, and diverse shapes.

3.1. Base Network. The Faster R-CNN model is a classical two-stage target detection network, which is now widely used in the industry. The model consists of the backbone network, region proposals network, ROI Pooling network, and fully connected network. The backbone network is used to extract the image feature maps and input the features into the next stage. The RPN generates and regresses the detection proposals based on the image size, which filter the useful proposals with the extracted feature maps. The ROI Pool network is applied to ensure the same feature size. The fully connected network classifies and performs bounding box regression on the features extracted from the ROIs pooling layer network. The predicted and true values are acquired to calculate the network loss for reverse gradient propagation. Finally, the Faster R-CNN model obtains the

classification and bounding box results of the predicted target by calculating the network layers of each part on the input image.

3.2. Improved Faster R-CNN. In this paper, an improved method based on the Faster R-CNN is proposed for detecting surface defects of ferrite shields, which consist of bidirectional feature fusion network, adaptive anchor boxes, ROI Align, and Soft nonmaximum suppression. The improved network model is shown in Figure 4. Firstly, the ResNet-50 network is used to extract features from the image, and the bidirectional fusion feature module is introduced to enhance the feature extraction capability. Then, the adaptive anchor boxes are adopted in the region proposals network to get the appropriate prediction frame. After that, ROI Align is used to extract the information in the feature map of the prediction frame which is captured in the previous step. Finally, the labels and positions of the target are predicted and regressed by a fully connected network incorporated with Soft nonmaximum suppression.

3.2.1. Feature Extraction. The VGG-16 is adopted as the backbone to extract image features, and the upsampling is used to enhance the representation ability and the receptive field [34] of the feature map. A larger receptive field means that larger targets can be detected, but the detection ability for smaller targets is therefore reduced. Thus, the traditional Faster R-CNN model is unable to detect the small defects effectively, since it loses a large amount of image feature information by upsampling to enlarge the receptive field of the predicted feature layer.

Therefore, we adopt ResNet-50 as the backbone network to extract image feature maps, which is able to obtain richer information from complex backgrounds and accelerate model training while maintaining detection speed. Besides, in order to improve the detection accuracy of the small target, we introduce a bidirectional feature fusion network into the Faster R-CNN. The specific structure of the bidirectional feature fusion network is as follows: the top-down

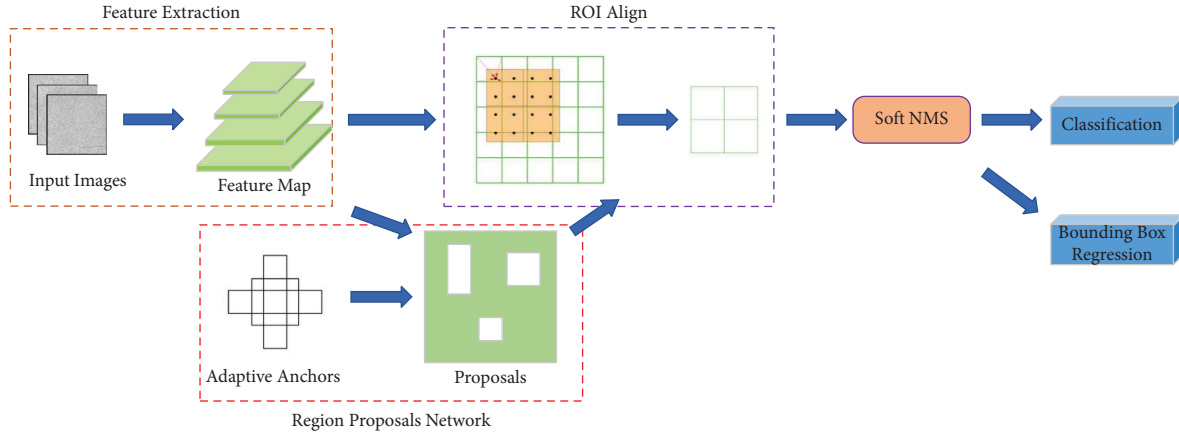


FIGURE 4: Overall network structure diagram.

feature pyramid is used to down-sample the higher level feature layer and fuse it with the lower level feature layer. Then, the top-down feature pyramid is used to up-sample the new lower level feature layer and fuse it with the new higher level feature mapping.

The high-level feature layer has larger receptive field and stronger semantic information, so the top-down feature pyramid is added to get the feature map with stronger semantic information, so as to improve the detection ability of the model for large targets. However, much low-level detailed information is lost after multiple upsampling and convolution operations. It contains a large number of feature information of small target defects, as well as positioning information that can improve the positioning accuracy. As shown in Figure 5, the low-level detailed features and positioning information is passed to high-feature layers by adding a bottom-up feature pyramid, which can retain semantic information at the same time. For ferrite shield surface defects, the adopted bidirectional feature fusion network is able to retain the localization information of defective flaws without losing their semantic information. The network is also improved to detect large crazing defect targets with high semantic information and small impurity defect targets with high localization information.

3.2.2. Region Proposal Network. The RPN stage is conducted by three steps. Firstly, a sliding window convolution operation is processed on the feature map extracted by the bidirectional feature fusion network, resulting in a feature map of 256 channels. Secondly, the predicted anchor boxes of different size and aspect ratio are generated within each anchor point. The anchor point refers to the rectangular center point of the original image, corresponding to each pixel in the feature map. Finally, the category of the predicted anchor box is calculated by predicting the obtained feature map. The bounding box of the defective target is obtained by boundary regression.

The conventional Faster R-CNN generates nine prediction anchor boxes of different sizes with aspect ratios of 2:1, 1:1, and 1:2 at the anchor points in the RPN stage, and then performs class determination and boundary regression

on the prediction anchor boxes. However, for ferrite shield surface defects of various shapes, the conventional aspect ratio prediction anchor boxes do not fit well for narrow crazing and small pit and impurity targets. Thus, we propose an adaptive method for generating anchor boxes which are shown in Figure 6.

As shown in Figure 6, the proposed adaptive method for generating anchor boxes consists of three steps. Firstly, the aspect data of targets in the dataset are extracted. Secondly, twelve true target boxes are randomly selected as cluster centers, and the intersection over union (IOU) between target boxes is considered when clustering twelve anchor boxes of different sizes. Finally, the obtained anchor frames are optimized with genetic algorithms. Therefore, in the RPN stage of the network in this paper, adaptive anchor boxes are generated to improve the ability of detecting narrow and fine defect targets on the surface of the ferrite shield.

3.2.3. ROI Align. ROI Pool is the basic segment to extract the features of candidate frames from the feature map. Firstly, the candidate frame boundaries are quantified into integer coordinates. Then, the quantified bounding frames are divided equally into an integer number of cells. After that, the boundaries of each cell are quantified, and finally the features in each cell are integrated and predicted. These two quantization processes in the ROI Pool would cause some displacement bias in the initially obtained candidate frames, thus affecting the detection results of the candidate frames. In order to eliminate the influence caused by this quantization step and improve the detection accuracy of shield surface defects, the ROI Pool in the Faster RCNN model is replaced with ROI Align.

ROI Align is a region feature aggregation method proposed by Kaiming He in Mask RCNN, as shown in Figure 7. The pixel value of the floating-point coordinate is obtained first by the bilinear interpolation method, followed by the max pooling or average pooling operation. This method eliminates the quantization operation, which causes numerical bias and converts the entire feature aggregation process into a continuous operation, thus reducing the displacement bias in the extraction of candidate frame features.

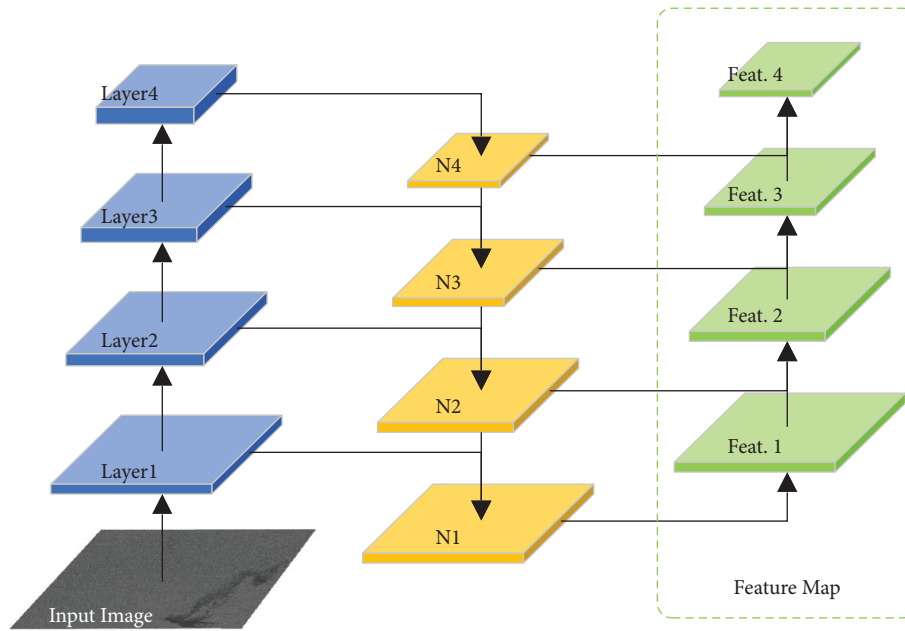


FIGURE 5: The structure of the bidirectional feature fusion network. The blue blocks (layers 1, 2, 3, and 4) are the feature layers extracted from ResNet-50. The yellow blocks (N1, 2, 3, and 4) are the new feature layers obtained from the top-down feature pyramid. The green blocks (feat.1, 2, 3, and 4) are the feature layers obtained from the bidirectional feature fusion network.

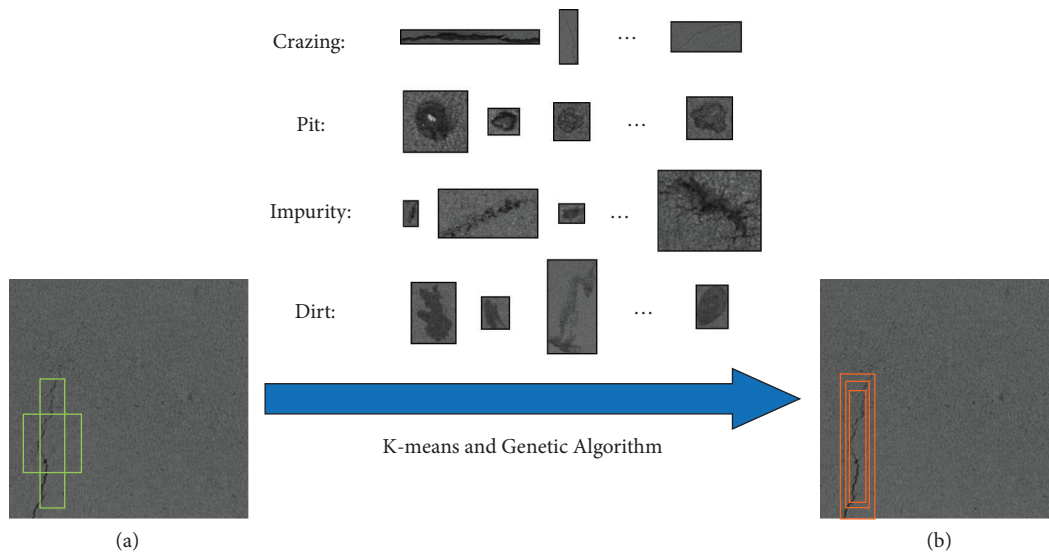


FIGURE 6: The flowchart of adaptive anchor boxes. (a) The anchor boxes adopted in the traditional Faster R-CNN. (b) The anchor boxes generated by *k*-means and genetic algorithm.

In this paper, the candidate frames and features obtained in the RPN are input into the ROI Align layer to obtain a 7×7 feature layer, and the features are input into the subsequent classification and regression network.

3.2.4. Soft NMS. After generating the final prediction feature layer obtained by the ROI Align layer, the classification results with boundary parameters are generated by the following fully connected network.

For object detection method, the nonmaximum suppression operation is usually used as the postprocessing

method to eliminate overlapping detection frames and reduce the false detection rate. The common approach is to sort the detected boxes by score, keep the boxes with the highest score, and then remove the other boxes with an overlap greater than the given threshold. However, this method may delete the overlapping target boxes by mistake, which actually exists. Therefore, the Soft non-maximum suppression is applied to filter target frames, which can reduce the score of the overlapped detection frames. It also can improve the prediction accuracy of overlapping targets. The IOU of the detection frame A and B is defined as follows:

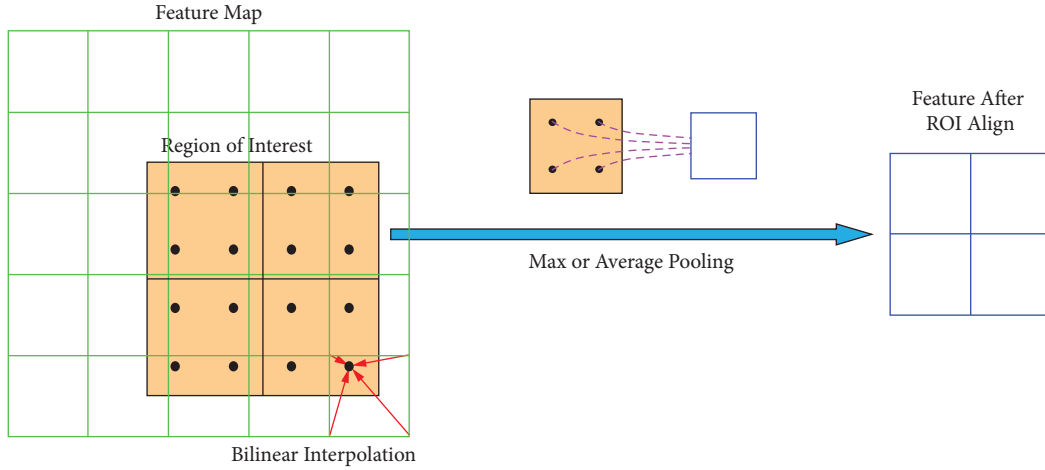


FIGURE 7: The flowchart of adaptive anchor. The values of the black points in the orange ROI are obtained by bilinear interpolation represented by the red arrows, and the purple feature map is obtained using the max or average pooling represented by the purple dashed line.

$$IOU = \frac{A \cap B}{A \cup B} = \frac{S_U}{S_A + S_B - S_U}, \quad (1)$$

where S_A and S_B are the areas of the detection frame A and B, respectively, and S_U is the overlapping area of the detection frames A and B.

The form of the traditional NMS and Soft NMS algorithms are as follows:

$$S_i = \begin{cases} S_i, IOU(M, b_i) < N_t, \\ 0, IOU(M, b_i) \geq N_t, \end{cases} \quad (2)$$

$$S_i = \begin{cases} S_i, IOU(M, b_i) < N_t, \\ S_i e^{-IOU(M, b_i)^2 / \sigma}, IOU(M, b_i) \geq N_t, \end{cases} \quad (3)$$

where formula (2) is the traditional NMS, formula (3) is Soft NMS. S_i is the score of the detection box i , M is the maximum score frame, b_i is the detection boxes other than the maximum score box, and N_t is the threshold set. The Soft nonmaximum suppression attenuates the score of detection frames whose IOU is greater than the given threshold, instead of setting the score to 0. This method can reduce false and missed detections of overlapping targets.

4. Experimental and Result Analysis

4.1. Training Platform and Parameter Settings. The methods in this article are implemented in the Pytorch framework in Windows 10. The computer is equipped with Intel(R) Core(TM) i9-10920X CPU and NVIDIA Ge-Force RTX 2080Ti GPU.

The self-built ferrite shield surface defects (FSSD) dataset is construed for training the proposed model. The random gradient descent (SGD) training model is selected. The initial learning rate is set to 0.01, the momentum is set to 0.9, and the weight is attenuated to 0.0005. The warm-up strategy is adopted to change the learning rate with the training process, that is,

from the initial value of 0.00058 to 0.01. We decay the learning rate with a cosine annealing as a way to increase the speed of model convergence, as shown in Figure 8. All layers are randomly initialized with Kaiming normal distribution. Due to the limitation of the GPU memory, the batch-size of the model training in this paper is set to 8 and the epoch of the training is set to 50. The models for each epoch are saved and the model with the best accuracy is selected as our final model.

The expressions of a cosine annealing method are as follows:

$$\eta_t = \eta_{\min} + \frac{1}{2} (\eta_{\max} - \eta_{\min}) \left(1 + \cos \left(\frac{T_{\text{cur}}}{T_{\text{max}}} \pi \right) \right), \quad (4)$$

where η_t is the learning rate of the t epoch, η_{\min} is the minimum learning rate, η_{\max} is the maximum learning rate, T_{cur} is the current epoch, and T_{max} is the total epoch.

4.2. Evaluation Metrics. In this paper, the mean average precision (mAP), which is the average accuracy of the model to detect each type of defects, is used as an evaluation metrics for the overall quality of the model training. A higher mAP represents a better overall ability of the model to detect defects. It is calculated as follows:

$$\begin{aligned} P &= \frac{TP}{TP + FP}, \\ R &= \frac{TP}{TP + FN}, \\ AP &= \int_0^1 P dR, \\ mAP &= \frac{1}{n_j \sum_{j=1}^{n_j} AP_j}, \end{aligned} \quad (5)$$

where TP denotes the true positives, FP the false positives, FN the false negatives, P the accuracy rate, and R the

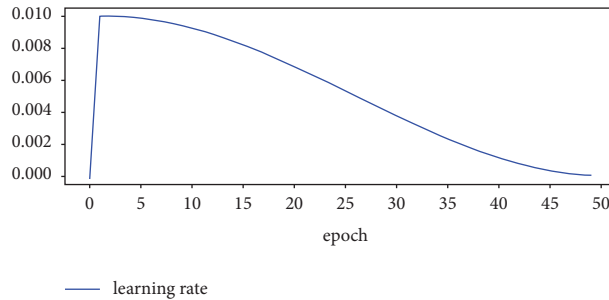


FIGURE 8: Learning rate graph.

recall rate. After calculating the precision and recall for a class in the whole self-built dataset, we consider the average precision (AP) as the area under the precision \times recall curve.

4.3. Analysis of Experimental Results. A visual assessment of the predictions for images in the test set was performed for the proposed method (Figure 9). In these figures, most of the four types of defects have been identified and positioned correctly due to the proposed tricks. The proposed method can not only detect large defects, but also can detect small defects. For example, both the small crazing in image 2 and the large crazing in image 4 are classified accurately. The impurity defects with relative small size are detected correctly in image 1, 5, 6, and 11. On the other hand, the proposed method can also handle the problem of various shapes. The crazing defects displayed in Figure 9 showing different shapes, and most of them are identified correctly. These effective detection benefits from the proposed bidirectional feature fusion and improved RPN, which can provide rich image feature and adaptive anchor boxes for small and varying shapes of defects. However, there was some confusion between pit and impurity at the time of classification, probably due to the proximity of their shapes in the image.

The performance achieved by the proposed model for four types of defects is presented in Table 3, from which we can see that the model has good detection performance for various types of defects on the surface. The mAP for the test set is up to 81% which is basically equal to the performance of the manual detection. Specifically, the AP value of the pit defect is the highest and reaches 0.926 among the four types of defects, probably due to the greater contrast between the defect and the background. The AP of dirt is the lowest, which is 0.674. We analysed that due to the small number of real samples of dirt defects in the dataset, the gamma and rotation transform do not greatly improve the fitness of the model to this defect.

4.4. Ablation Experiments. We performed ablation experiments to demonstrate the usefulness of each module in the proposed model. This model was compared with different improved models on the test set of the FSSD dataset to validate its effectiveness. The specific six improved models are as follows:

- (1) Model 1 is the original Faster R-CNN model
- (2) Model 2 uses ResNet-50 as the feature extraction network of Faster R-CNN model
- (3) Model 3 introduces a bidirectional feature fusion network on model 2
- (4) Model 4 adds adaptive anchor boxes in the RPN stage on the basis of model 3
- (5) Model 5 replaces ROI Pool with ROI Align on the basis of model 4
- (6) Model 6 adopts Soft nonmaximum suppression instead of nonmaximum suppression on the basis of model 5

The test results of the six experimental models are shown in Figure 10. The experimental results show that model 1 can only detect individual obvious defects, resulting in a large number of missed detections. Model 2 also has missed detections and does not identify crazing defects well. Model 3 can detect most of the defects, but there are false detections and missing detections, and the defect location is inaccurate. Model 4, 5, and 6 can almost detect all the defects and these three models have better detection accuracy for small and slender defects due to the adaptive anchor module. Model 5 has added ROI Align to obtain more accurate location information. Model 6 uses Soft NMS instead of NMS, so that the target score of nonmaximum fraction defects is less suppressed and the detection of overlapping defects is more accurate.

The specific results of the experiments are shown in Table 4. We can find that the mAP of model 1, which is the original Faster R-CNN model, is 67.1%. After replacing the original feature extraction network with ResNet-50, the detection accuracy for both pit and impurity of model 2 are improved significantly. The mean average accuracy of model 3 is 9.5% higher than that of model 1, in which the accuracy of detection for each type of defect has been improved. Model 4 adds adaptive anchor boxes, which is more adaptable to the shape of various categories of defects. Although the accuracy of detection for crazing defects is reduced, the accuracy of detection for pit and impurity has improved by about 1%. Model 5 uses ROI Align to replace the ROI Pool in the original network, which fit the location of defects more accurately and enhance the detection accuracy for pit and dirt. While the accuracy for impurity is lower than model 2 due to the lack of feature fusion, the

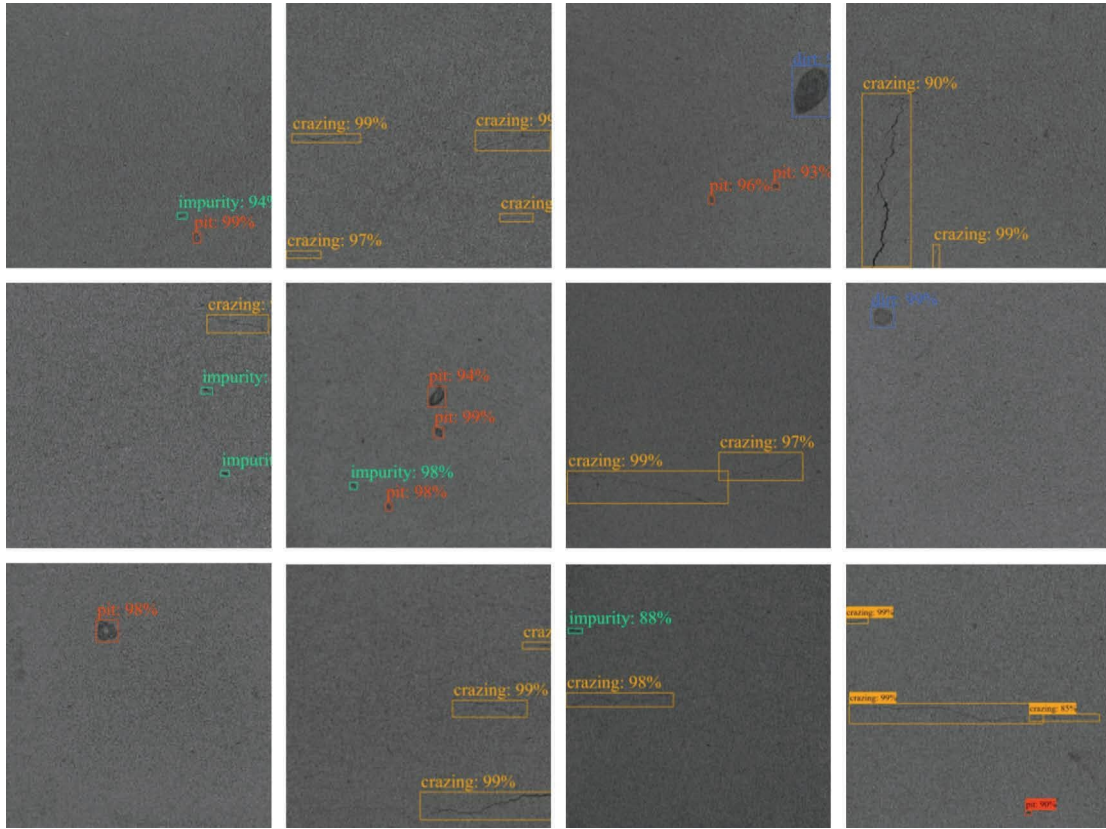


FIGURE 9: The detection results of the proposed model.

TABLE 3: The experimental data of the algorithm on all defects.

Defects	Crazing	Pit	Impurity	Dirt
AP	0.853	0.926	0.790	0.674
mAP	0.810			

proposed model 6 is better than other models in detecting crazing, pit, and dirt. In addition, the average accuracy of all the defects is improved and the mean average accuracy has increased by 13.9% compared with the original Faster R-CNN model.

As shown in Figure 10 and Table 4, it can be seen that the original Faster R-CNN model is not effective for the detection of shield defects. In this paper, after introducing ResNet-50, bidirectional feature fusion, the model has enhanced the extraction ability of fine targets, such as pit and impurity, on weak information features. Then, the addition of adaptive anchor boxes and ROI Align is helpful for the fitting accuracy of defects with diverse shapes. Finally, Soft NMS improves the ability of the model to defect overlapping targets. Therefore, the model proposed in this paper has excellent detection performance for ferrite shield surface defects.

4.5. Algorithm Comparison. The proposed model in this paper is compared with other mainstream model to further verify the advantages of our model for ferrite shield surface

defect detection. Table 5 shows the detection performance and the mAP of different model for crazing, pit, impurity, and dirt defects. The results show that the proposed model has a great improvement in the detection accuracy of crazing, pit, and impurity compared with other mainstream algorithms. The overall mAP is 15.3%, 12.9%, 11.0%, 8.2%, and 0.3% higher than that of SSD, YOLOv3, Retina-Net [35], Cascade R-CNN [36], and YOLOv5, respectively. Since the multistage detector reduces the effect of quality mismatch in the augmented data, our model is about 4% less accurate than Cascade R-CNN for the detection of dirt. For YOLOv5, because of the mosaic image enhancement, its ability to detect dirt defects with fewer samples is better than the present model. Although the mAP of YOLOv5 is similar to the proposed model, and the detection speed is faster than the proposed model, we found that it has many missed detections for large crazing and small target defects (pit and impurity). The reason is that the YOLOv5 model has fewer parameters and a shorter network depth than the proposed model, so it is slightly less powerful than our model in detecting crazing with less feature information and small targets of pit and impurity. The recall of the proposed model

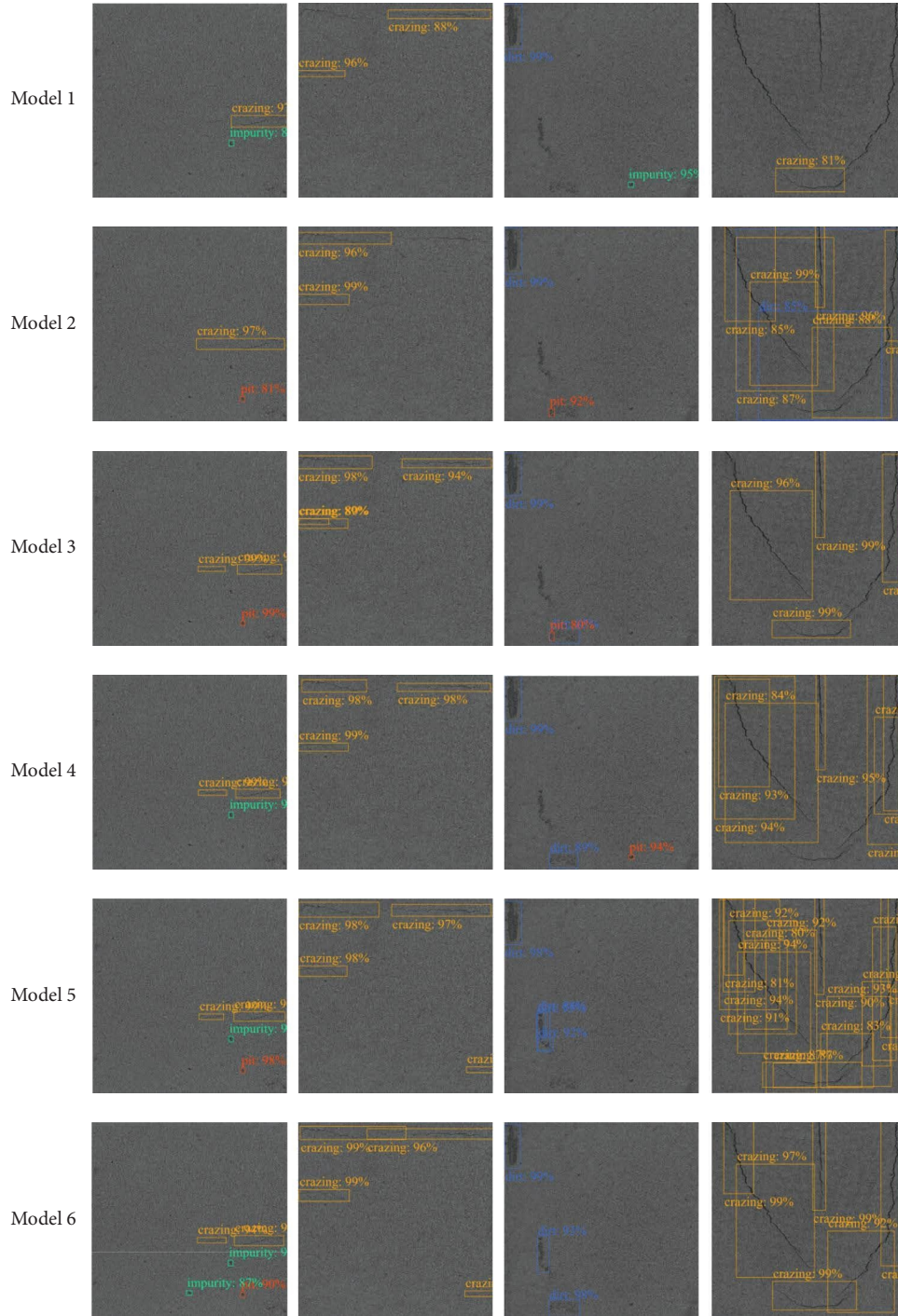


FIGURE 10: The detection results of different models.

TABLE 4: The data results comparison of the six models.

Models	Crazing	Pit	Impurity	Dirt	mAP	mAP increasing
1	0.771	0.694	0.637	0.579	0.671	—
2	0.731	0.858	0.854	0.489	0.734	0.063
3	0.845	0.896	0.770	0.551	0.766	0.095
4	0.820	0.900	0.780	0.592	0.773	0.102
5	0.791	0.924	0.809	0.643	0.792	0.121
6	0.853	0.926	0.790	0.674	0.810	0.139

TABLE 5: The model of SSD, YOLOv3, Retina-Net, and Cascade R-CNN are compared with the experimental results of the proposed model.

Models	Crazing	Pit	Impurity	Dirt	mAP	Time (ms)
SSD	0.626	0.817	0.645	0.539	0.657	39
YOLOv3	0.625	0.818	0.686	0.593	0.681	22
Retina-Net	0.689	0.779	0.674	0.659	0.700	76
Cascade R-CNN	0.745	0.759	0.691	0.718	0.728	82
YOLOv5	0.744	0.845	0.784	0.855	0.807	16
Proposed	0.853	0.926	0.790	0.674	0.810	51

is much higher than YOLOv5, and it is a better choice for industrial detection with our proposed more stable model.

5. Conclusions

An effective detection method for the surface defects of ferrite shields is proposed, which is based on the improved Faster RCNN. Firstly, ResNet-50 is adopted as the backbone to extract rich surface image feature, followed by the bidirectional feature fusion network layer to obtain the accurate location information and rich semantic features at the same time. Then, the RPN improved by k -means clustering and genetic algorithm is demonstrated to select anchor boxes for defects with various sizes and shapes adaptively. Moreover, ROI Align is applied to eliminate the candidate frame position bias and improve the defect localization accuracy. Soft NMS is used to increase the ability of the model to detect overlapping targets. The experimental results show that the average accuracy of the proposed method for crazing, pit, impurity, and dirt is 85.3%, 92.6%, 79.0%, and 67.4%, respectively. Importantly, the mean average precision is 81.0% which is better than SSD, YOLO, Retina-Net, and Cascade RCNN. In conclusion, an effective method to solve the problem of small and various defects detection is presented, which can provide an alternative for automatic ferrite shield detection. But, there is still a certain gap in the detection speed between the proposed method and the one-stage model. The mAP needs to be further increased to improve the practicability of the algorithm. In the future, we aim to accelerate the real-time performance of the model by model pruning or knowledge distillation, and construed new model to increase the mAP.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 52075152) and Collaborative

Innovation Center of Intelligent Green Manufacturing Technology and Equipment, Shandong (IGSD-2020-006). Also, this work was supported by the Key R&D Projects in Hubei Province, Hubei (2022BBA0016).

References

- [1] H. Fan, Y. Hou, and B. Li, "Visual detection of defective fasteners based on area features," *Journal of the China Railway Society*, vol. 43, no. 8, 2021.
- [2] Q. Wang and C. Zhang, "Research on defect detection technology of automotive wheel casting with pattern recognition," *Foundry Technology*, vol. 38, no. 12, p. 4, 2017.
- [3] Y. Yuan Xiao-cui, W. Lu-shen, and C. Hua-wei, "Rail image segmentation based on Otsu threshold method," *Optics and Precision Engineering*, vol. 24, no. 7, pp. 1772–1781, 2016.
- [4] Z. Yang and X. Dai, "Detection and implementation of food packaging printing defects based on image alignment," *Computer Science*, vol. 42, no. 8, p. 4, 2015.
- [5] S. Mei, Y. Wang, and G. Wen, "Automatic fabric defect detection with a multi-scale convolutional denoising autoencoder network model," *Sensors*, vol. 18, no. 4, p. 1064, 2018.
- [6] X. Tao, D. Zhang, W. Ma, X. Liu, and D. Xu, "Automatic metallic surface defect detection and recognition with convolutional neural networks," *Applied Sciences*, vol. 8, no. 9, p. 1575, 2018.
- [7] L. Yang, Z. Wang, and S. Gao, "Pipeline magnetic flux leakage image detection algorithm based on multiscale SSD network," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 1, pp. 501–509, 2020.
- [8] V. F. Fadli and I. O. Herlistiono, "Steel surface defect detection using deep learning," *International Journal of Innovative Science and Research Technology*, vol. 5, no. 7, pp. 244–250, 2020.
- [9] T. Wang, Y. Chen, M. Qiao, and H. Snoussi, "A fast and robust convolutional neural network-based defect detection model in product quality control," *International Journal of Advanced Manufacturing Technology*, vol. 94, no. 9–12, pp. 3465–3471, 2018.
- [10] Y. He, K. Song, Q. Meng, and Y. Yan, "An end-to-end steel surface defect detection approach via fusing multiple hierarchical features," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 4, pp. 1493–1504, 2020.
- [11] Q. Ren, J. Geng, and J. Li, "Slighter Faster R-CNN for real-time detection of steel strip surface defects," in *Proceedings of the Chinese Automation Congress*, pp. 2173–2178, IEEE, Xi'an, China, November 2018.
- [12] 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, pp. 1–13, 2021.
- [13] B. Hu and J. Wang, "Detection of PCB surface defects with improved faster-RCNN and feature pyramid network," *IEEE Access*, vol. 8, Article ID 108335, 2020.
- [14] N. Zeng, P. Wu, Z. Wang, H. Li, W. Liu, and X. Liu, "A small-sized object detection oriented multi-scale feature fusion approach with application to defect detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–14, 2022.
- [15] J. Xing and M. Jia, "A convolutional neural network-based method for workpiece surface defect detection," *Measurement*, vol. 176, Article ID 109185, 2021.

- [16] H. Gao, Y. Zhang, W. Lv, J. Yin, T. Qasim, and D. Wang, "A deep convolutional generative adversarial networks-based method for defect detection in small sample industrial parts images," *Applied Sciences*, vol. 12, no. 13, p. 6569, 2022.
- [17] H. Huang, X. Tang, F. Wen, and X. Jin, "Small object detection method with shallow feature fusion network for chip surface defect detection," *Scientific Reports*, vol. 12, pp. 3914–3919, 2022.
- [18] L. Zhu, C. Jia, J. Zhang, L. Wang, and Y. Jian, "YOLO-C: a new surface defect detection model of steel plate based on YOLO optimisation," *International Journal of Wireless and Mobile Computing*, vol. 23, no. 1, pp. 46–56, 2022.
- [19] 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*, pp. 770–778, IEEE, Las Vegas, NV, USA, June 2016.
- [20] K. Wang, J. Liew, Y. Zou, D. Zhou, and J. Feng, "Panet: few-shot image semantic segmentation with prototype alignment," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9197–9206, IEEE, Los Alamitos, CA, USA, 2019.
- [21] Q. Zeng and G. Zhang, "Adaptive K-means clustering algorithm based on distance threshold," *Journal of Zhengzhou University*, vol. 48, no. 4, pp. 90–94, 2016.
- [22] Y. Ma and W. Yun, "Research progress of genetic algorithm," *Application Research of Computers*, vol. 29, no. 4, pp. 1201–1206, 2012.
- [23] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969, IEEE, Venice, Italy, October 2017.
- [24] N. Bodla, B. Singh, R. Chellappa, and L. S. Davis, "Soft-NMS--improving object detection with one line of code," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 5561–5569, IEEE, Venice, Italy, 2017.
- [25] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
- [26] Z. Luo and Z. Wang, "Image processing and application based on OpenCV technology," *Strait Technology and Industry*, vol. 9, p. 5, 2019.
- [27] B. O. Koopman, "The theory of search. II. Target detection," *Operations Research*, vol. 4, no. 5, pp. 503–531, 1956.
- [28] H. Yin, B. Chen, and Y. Chai, "A review of vision-based target detection and tracking," *Journal of Automation*, vol. 42, no. 10, pp. 1466–1486, 2016.
- [29] W. Liu, A. Dragomir, E. Dumitru et al., "Ssd: single shot multibox detector," in *Proceedings of the European Conference on Computer Vision* Springer, Amsterdam, Netherlands, 2016.
- [30] J. Redmon and A. Farhadi, "Yolov3: an incremental improvement," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1804–2767, IEEE, Salt Lake City, UT, USA, April 2018.
- [31] R. Girshick, J. Donahue, and T. Darrell, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, IEEE, Columbus, OH, USA, June 2014.
- [32] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440–1448, IEEE, Santiago, Chile, December 2015.
- [33] S. Ren, K. He, and R. Girshick, "Faster r-cnn: towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, pp. 91–99, 2015.
- [34] W. Luo, Y. Li, and R. Urtasun, "Understanding the effective receptive field in deep convolutional neural networks," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pp. 4905–4913, Springer, Berlin, German, December 2016.
- [35] T. Lin, P. Goyal, and R. Girshick, "Focal loss for dense object detection," in *IEEE International Conference on Computer Vision*, vol. 42, pp. 2980–2988, IEEE, 2015.
- [36] Z. Cai and N. Vasconcelos, "Cascade r-cnn: delving into high quality object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6154–6162, IEEE, Salt Lake City, UT, USA, 2018.