WILEY | Hindawi

Research Article

Research on Surface Defect Detection of Rare-Earth Magnetic Materials Based on Improved SSD

Bin Zhang ⓘ**, Shuqi Fang** ⓘ**, and Zhixi Li** ⓘ

School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China

Correspondence should be addressed to Shuqi Fang; fsq1611@sina.com

In order to overcome the limitation of manual visual inspection of surface defects of rare-earth magnetic materials and increase production efficiency of traditional rare-earth enterprises, a detection method based on improved SSD (Single Shot Detector) is proposed. The SSD model is improved from two aspects for better performance in the detection of small defects. First of all, the multiscale receptive field module is embedded into the backbone network of the algorithm to improve the feature extraction ability of the model. Secondly, the interlayer feature fusion strategy of bidirectional feature pyramid in PANet (path aggregation network) is integrated into the model. In order to enhance the detection ability of the model, the high-level semantic information is strengthened by an efficient channel attention mechanism. The detection speed of the improved SSD algorithm is 55FPS, and the mAP (mean Average Precision) is up to 83.65%, which is 3.41% higher than of the original SSD algorithm, and the ability to identify small defects is significantly improved.

1. Introduction

Computer vision technology has very important application and theoretical significance in defect detection of rare-earth magnetic materials. At present, most magnetic material manufacturers in China use the traditional method of manual sorting to classify their products. It is not only time consuming and laborious, but also increasing the cost of products. At the same time, an overreliance on manual sorting raises a problem that cannot be ignored: operators in the sorting process will inevitably cause visual and physical fatigue. Lack of attention for a long time will directly lead to an increase in the rate of false and missed detection. With the development of computer technology and image processing technology, automatic detection based on image processing technology is an inevitable trend. The visual-based surface quality inspection method has very important research value, for example, the steel surface damage detection [1], the railway track defect detection [2], the wafer electron microscope image defect detection [3], and a wide range of applications in other fields [4, 5].

In previous studies, researchers usually use traditional machine learning methods to carry out a series of studies in the field of defect recognition. Chang et al. [6] proposed a defect detection method based on LVQ neural network. Aiming at recognizing the defects on the LED wafer, the geometric features and texture features are extracted by analyzing each ROI for detection. Yazdchi et al. [7] presented a texture segmentation technology based on multifractal dimensions to detect steel surface defects. Fourier analysis in the time domain is utilized to detach the defective region from the image and specify its position. The features of multifractal dimension, mean and variance of column variance, and maximum value of principal component vector are extracted as the input of the three-layer and multilayer perceptron classifier. Demetgul et al. introduced a KNN-based fabric defect classification algorithm, which uses wavelet transform, threshold, and morphology for image processing [8]. Jeon et al. proposed a method based on wavelet reconstruction to detect cracks at the corners of billets [9]. Sun et al. proposed a new single-shot target detection network with a mask prediction branch. They proposed an improved RFB module to increase the size of receptive field for better performance on small object detection [10]. Although these methods have achieved good results, they usually require explicit feature extraction, which leads to unsatisfactory generalization of detection methods.

In recent years, with the development of deep learning, its outstanding performance in the field of image processing has attracted people's attention. Deep learning avoids the manual extraction of features and realizes automatic feature extraction. Combined with CNN (Convolutional Neural Network), relevant researchers have carried out a series of research work in the field of defect detection of motor magnetic tiles [11–13]. Due to the various types of rare-earth magnets, there is currently little work on the automatic detection of rectangular rare-earth magnetic patches domestically and overseas. Therefore, this research combines machine vision and deep learning and other related theories and adopts SSD (Single Shot Detector) [14] as the prototype of defective target detection and improves it. The improved SSD algorithm has a better recognition effect for small defects, and the detection accuracy is significantly improved compared with the original SSD algorithm.

2. SSD Target Detection Algorithm Model

The industrial defect detection task requires high detection speed. Compared with two-stage model such as Faster R-CNN [15], the one-stage target detection algorithm is more suitable for actual production and practice. This research selects SSD as the basic network and improves it. The improved SSD can detect defects on the surface of the magnet more accurately and has a higher recognition rate for small targets. It provides a feasible method for the industrial scene where defect detection is needed.

2.1. SSD Algorithm Principle. SSD is a one-stage target detection algorithm that borrows ideas from the anchors mechanism in Faster R-CNN and uses multiscale feature map for detection. The backbone feature extraction network of the SSD algorithm is the classic VGG-16 [16] convolutional neural network, which has been modified. The SSD algorithm replaces fc6 and fc7 in the VGG-16 with a convolutional layer structure and, on this basis, adds four additional convolutional layers to obtain more feature maps for detection. The structure of the SSD network is shown in Figure 1.

The feature extraction network uses feature maps of different sizes output by the specific effective feature layers for prediction and outputs the location information of the target and the confidence of the category. Among them, the output feature map scale of the Conv4_3 layer structure is 38×38, the output feature map scale of the fc7 layer structure is 19×19, and the output feature map scale of the Conv6_2 layer structure is 10×10. The output sizes of Conv7_2, Conv8_2, and Conv9_2 layers are 5×5, 3×3, and 1×1, respectively. Six feature maps of different scales can detect targets of different sizes. The front feature map has a smaller receptive field and is mainly used to identify small targets, while the back feature map has a larger receptive field and is used to detect large targets. Each pixel of the feature map has several default boxes, as shown in Figure 2. The proportion between the size of the default box and the picture can be expressed as follows:

$$s_k = s_{\min} + \frac{s_{\max} - s_{\min}}{m - 1} (k - 1), \quad k \in [1, m]. \tag{1}$$

Among them, m is the number of valid feature map (excluding Conv4_3). $S_{\max}$ and $S_{\min}$ represent maximum and minimum proportions, with values of 0.9 and 0.2, respectively. The default box scale of the first effective feature layer is independently set to 0.1, and the scale is 30 on the 300×300 original image. The aspect ratio of the default box is set as $\alpha_r = \{1, 2, 3, 1/2, 1/3\}$; then, the width and height of the default box can be obtained by the following formula:

$$w_k = s_k \sqrt{\alpha_r},$$
$$h_k = \frac{s_k}{\sqrt{\alpha_r}}. \tag{2}$$

It is special when $\alpha_r = 1$. In addition to the default box with the scale of s_k, there is another scale of s_k', and its calculation formula is as follows:

$$s_k' = \sqrt{s_k s_{k+1}}. \tag{3}$$

Since the feature map output by Conv4_3 is relatively special, the default boxes with an aspect ratio of 3 and 1/3 are not used. From this, the number of default boxes for each pixel of the feature layer as the center is 4, and the remaining layers are 6. So the number of SSD default boxes is 8732. Due to the large number of default boxes obtained and the limited number of GT (Ground Truth) matched by IOU value, the SSD algorithm uses hard negative mining to sample negative samples. Those with large errors are taken as negative samples, while those with small errors are considered as positive samples. The final sample ratio of positive and negative ones is controlled at about $1 : 3$.

The bounding box of SSD algorithm is fine-tuned by the default box, and its essence is a regression task. The offset value of each bounding box to the default box is inferred, and the final target position information is obtained through a transformation. The transformation process consists of two parts: encoding and decoding [14]. The position of the default box is represented by $d_b = (d_b^{cx}, d_b^{cy}, d_b^w, d_b^h)$, which respectively correspond to the coordinate of the center point and the width and height of the default box. The position of the bounding box is expressed as $b_b = (d_b^{cx}, d_b^{cy}, d_b^w, d_b^h)$, which respectively correspond to the coordinates of the center point and the width and height of the bounding box.

2.2. Problems with SSD Algorithm. Compared with Faster R-CNN and YOLO [17] algorithm, SSD algorithm has higher detection accuracy and detection speed, which benefits from its good design ideas. The SSD algorithm directly acts on the output information of the effective feature layer at different scales on the detection layer to generate the bounding box and the confidence of the detection target. The small target is detected by the shallow structure such as Conv4_3, and the large target is detected by the top effective feature layer such as Conv8_2.

The SSD algorithm uses a pyramidal sampling structure to express the semantic information of the image. The shallow
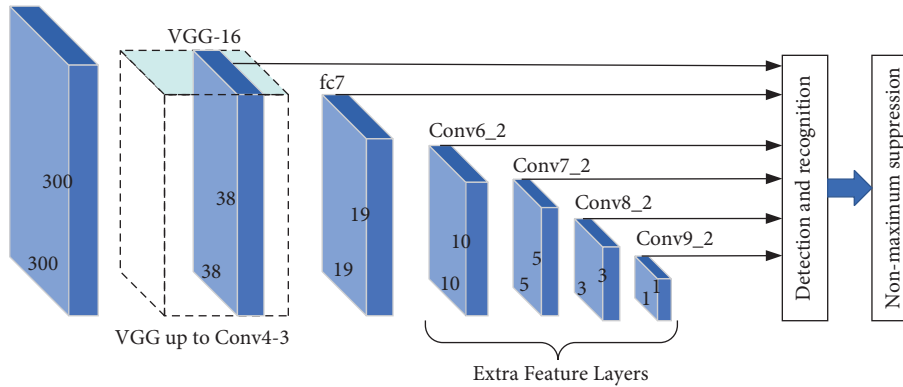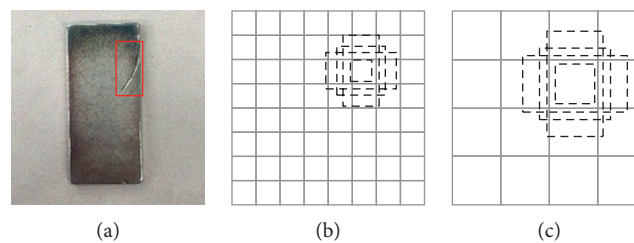
FIGURE 1: SSD algorithm structure.

(a)   (b)   (c)

FIGURE 2: Comparison of the default box and the real box under different scale feature graphs. (a) Image with GT boxes. (b) 8×8 feature map. (c) 4×4 feature map.

information has high pixels and has a strong ability to locate targets. However, the small target features obtained by the low-level convolutional layer lack semantic information. The nonlinear degree of the features is insufficient, and the features of their scales are not merged, resulting in the loss of part of the detailed information of the model learning. This will not be able to effectively use the contextual information related to the model, which is not conducive to the detection of related targets on the image. These deficiencies will lead to the failure to identify small-size defects in the defect detection process of rare-earth magnetic materials.

After multiple convolutions, the high-level feature maps for detecting large targets have larger receptive fields and the learned information is more abstract, but its lower resolution may still cause missed targets. The high-level semantic information in the SSD algorithm has not been enhanced, and there is still room for improvement.

3. Defect Detection Method of Rare-Earth Magnetic Patch Based on Improved SSD

3.1. PANet Bidirectional Feature Pyramid Structure. FPN (Feature Pyramid Network) [18] is a network topology structure that gathers high-level features and low-level features. It is different from the pyramid hierarchical sampling structure in SSD algorithm, so that the information learned by the model not only retains the location information but also contains stronger semantic information. Its network topology is shown in Figure 3. According to the topological structure, FPN performs upsampling (such as

bilinear interpolation and deconvolution) on the high-level feature map and predicts the low-level semantic information after horizontal superposition, so that the feature image has strong semantic information and achieves the effect of feature fusion.

Inspired by FPN, Shu Liu et al. proposed a path aggregation network called PANet; it was first applied to instance segmentation tasks and achieved excellent results [19]. Its framework is shown in Figure 4. It can be seen from the figure that PANet has improved the FPN structure, adopting the structure of the top-down and bottom-up bidirectional feature pyramid fusion and combining the downsampling on the basis of the original top-down horizontal superposition of FPN, adding the bottom-up path enhancement strategy. Its purpose is to shorten the information path and use the accurate positioning of the shallow level. This effectively improve the utilization rate of the underlying characteristics of the network.

3.2. Parallel Multiscale Convolutional Layer. At present, most of the algorithms with excellent detection results rely on backbone networks with strong feature extraction capabilities, such as the improved DSSD algorithm based on SSD [14]. This network replaces the original VGG-16 feature extraction network with a deeper network like ResNet101 and adds a DSSD network layer on this basis. These changes significantly improve the accuracy of the model, but require a lot of computational cost, and the detection speed is much lower than that of SSD. Some SSD series detection algorithms based on lightweight backbone networks, such as
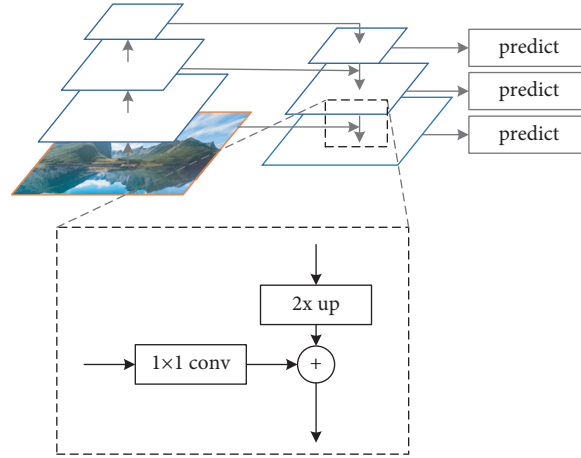
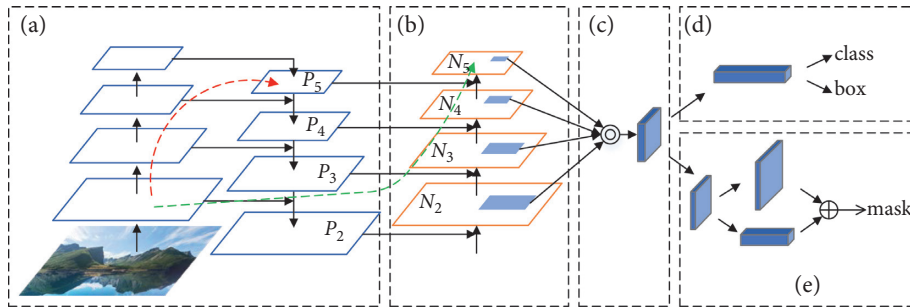FIGURE 3: FPN structure.

FIGURE 4: PANet structure.

MobileNet-SSD model [20], have relatively low detection accuracy despite their fast computing speed.

In order to avoid a large number of operations caused by deep convolutional neural network, Liu et al. [21] added a receptive field module to the top of VGG-16 network of SSD algorithm to obtain performance gains and control the computational cost within a controllable range. The structure of the receptive field module is similar to that of InceptionV4, and multiscale information is obtained by using parallel convolution kernels with different receptive fields. In the receptive field module, according to the size of the receptive field, each convolution branch performs a dilated convolution with different expansion coefficients to simulate the group receptive field mechanism in human vision. It also uses the residual learning idea and adds jump connections. This paper uses the receptive field module to enhance the VGG-16 backbone network and replaces the 5×5 convolutional layer in the receptive field module with two consecutively stacked 3×3 convolutional layers to achieve the same receptive field and reduce network parameters, as shown in Figure 5.

3.3. Efficient Channel Attention Mechanism. In the reasoning process of the SSD algorithm, the identification of medium and large targets is completed by the high-level feature map. Due to its low output resolution, the target may also be missed. In this paper, an efficient channel attention module ECA [22] is introduced to optimize the high-level and low-

resolution semantic information graph output by SSD algorithm. The obtained information graph with low resolution and high-semantic features can better identify medium and large defect targets, so as to achieve the purpose of improving mAP (mean Average Precision).

The attention mechanism is to imitate the signal processing mechanism of the human brain. This strategy has good adaptability and enhancement for computer vision tasks. ECA structure also obtains the importance degree of different characteristic channels through supervised learning. The nonlinear full-connection layer in SENet [23] is improved to avoid the impact of dimensionality reduction on the attention of the learning channel and ensure the efficiency and computing effect of the network. This module can achieve cross-channel interaction without dimensionality reduction. That is, after global average pooling of the output feature graph, each channel interacts with its K neighbors through rapid one-dimensional convolution. The topology of ECA module is shown in Figure 6.

The correlation of feature channels can be expressed as follows:

$$\omega = \sigma\left(\mathrm{CID}_k(y)\right), \tag{4}$$

where CID represents one-dimensional convolution operation, k represents the coverage of cross-channel interaction, and its size is proportional to the number of channel dimensions. When the channel dimension C is constant, the value of k can be determined by the following formula:
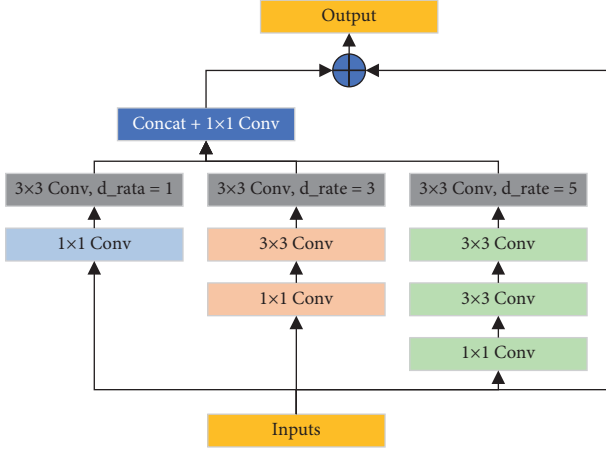
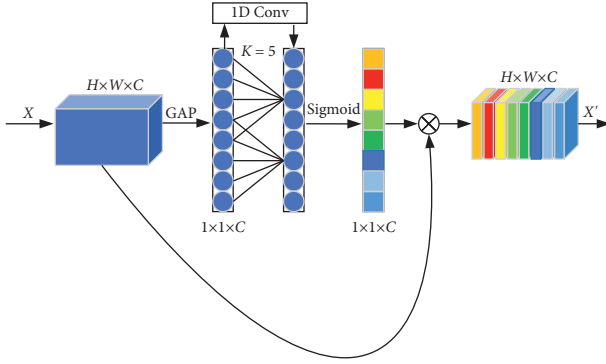FIGURE 5: Structure of the multiscale receptive field layer.

FIGURE 6: Structure diagram of the efficient channel attention module.

$$k = \psi(C) = \left| \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{odd}. \tag{5}$$

Here, odd is the closest odd number and γ and b are set to 2 and 1, respectively.

3.4. Establishment of the Improved SSD Network Model.

In order to make the original SSD algorithm better solve the problem of small-size defect identification and improve recognition effect, the SSD algorithm is improved based on the relevant theoretical basis analyzed in the above section. The framework of the defect detection method for rare-earth magnetic materials based on the improved SSD model is shown in Figure 7.

It can be seen from the figure that the improved SSD algorithm replaces the Conv6 and Conv7 convolutional layers of the original backbone extraction network by embedding multiscale receptive field modules Block_1, Block_2, and Block_3, thereby improving the ability of the VGG-16 backbone network to extract features and not affecting the inference speed of the detection algorithm. Conv8 and Conv9 layers are not modified. The effective feature layer dimension information output by the embedded Block_1 receptive field module is $19 \times 19 \times 1024$,

which has the same size as the output of the fc7 layer and can extract more effective feature information. Therefore, this study extracts the features of the convolutional layer to replace the output of the fc7 layer and uses the PANet-based bidirectional feature pyramid fusion with the Con4_3 and Con3_1 layer structure to make full use of the context information between different levels. The final output is 38×38 and 19×19 effective feature layers. The network structure of feature fusion between layers is shown in Figure 8. Among them, Block_1 adopts spatial pyramid pooling to increase the receptive field. After the interlayer fusion between the top-level feature layer and the low-level feature layer, alternate convolution of 1×1 and 3×3 convolutional layers is helpful to reduce the amount of parameters and extract very effective features. The deep effective feature layers Block_2, Block_3, Conv8_2, and Conv9_2 output four sizes of low-resolution and high-semantic information maps through the ECA channel attention mechanism network.

4. Rare-Earth Magnetic Patch Data Set and Experimental Analysis

4.1. Data Set. The image data sets used in this paper were all collected from Guangxi Jinyuan Rare Earth Co., Ltd. (Guangxi Nonferrous Metals Group). The data sets were manually classified with the assistance of sorting workers and photographed on site with industrial cameras. The focus of the research in this paper is four types of common defect data samples with category and location information, including unfilled corner, line mark, deformation, and crack, as shown in Figure 9. The defect targets are marked using LabeIImg labeling software before the experiment. A total of 1534 data samples were labeled and divided into training set, verification set, and test set in a ratio of $8:1:1$. After image annotation, XML files including image number, category, location of target marker box, and other information will be obtained for model training.

4.2. Network Model Training and Experimental Parameter Setting. The loss function of the improved SSD algorithm consists of two parts, namely the classification confidence loss function (Conf) associated with the target category and the location loss function (Loc) associated with the bounding box regression, as shown in

$$L(x, c, l, g) = \frac{1}{N} \left(L_{Conf}(x, c) \right) + \alpha L_{Loc}(x, l, g). \tag{6}$$

Among them, x represents the matching result; c and l respectively represent the confidence of the classification and the position of the bounding box; α is the scale factor, which is used to adjust the ratio of the bit loss function to the confidence loss function; g is the true box label; N represents the number of default boxes that match the actual box.

The confidence loss function can be expressed as follows:

$$L_{Conf}(x, c) = -\sum_{i \in Pos}^{N} x_{ij}^{p} \log(\hat{c}_i^{p}) - \sum_{i \in Neg} \log(\hat{c}_i^{0}). \tag{7}$$
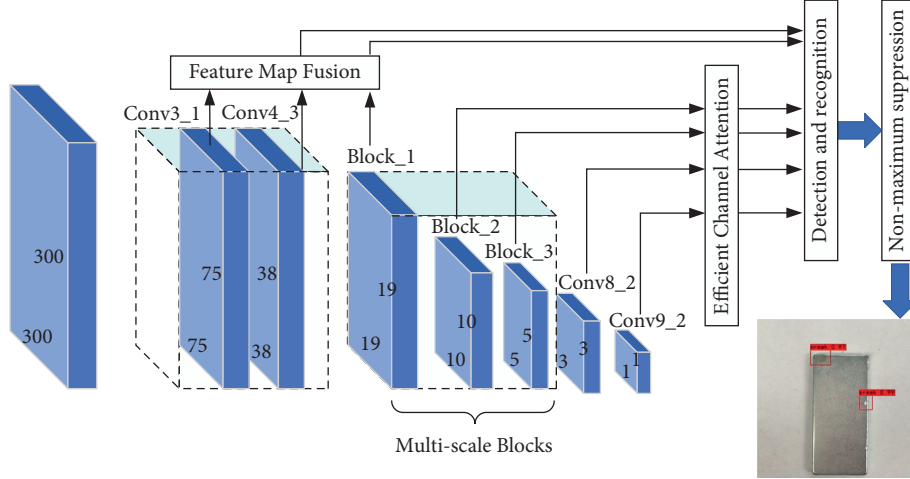
FIGURE 7: Improved SSD algorithm detection framework.

Among them, Pos represents the number of positive samples in the bounding box, Neg represents the number of negative samples in the bounding box, j represents the j-th true box, i represents the i-th bounding box, and x_{ij}^p represents whether it belongs to the p category. c_i^p represents the confidence of the corresponding category of the i-th bounding box, and $\widehat{c}_i^p$ represents the confidence of the background, which is calculated by the following formula:

$$\widehat{c}_i^p = \frac{\exp\left(c_i^p\right)}{\sum_p \exp\left(c_i^p\right)}. \tag{8}$$

The location loss function can be expressed as follows:

$$L_{\text{Loc}}(x, l, g) = \sum_{i \in Pos}^{N} \sum_{m \in \{cx, cy, w, h\}} x_{ij}^k \text{smooth}_{L1}\left(l_i^m - g_j^m\right). \tag{9}$$

Among them, k represents the category and $\text{smooth}_{L1}(x)$ represents the smooth L_1 norm,

$$\text{smooth}_{L1}(x) = \begin{cases} 0.5x^2, & |x| < 1, \\ |x| - 0.5, & \text{otherwise.} \end{cases} \tag{10}$$

All experiments in this paper are based on the Ubuntu 18.04 operating system and the NVIDIA high-speed hardware computing platform with $1 \times$ GeForce RTX 2080TI graphics card. The improved SSD target detection algorithm is implemented by the deep learning framework PyTorch 1.5.1 and python 3.6. The image data are all downsampled to 300×300 resolution as the input of the network. The parameters of this model are set as follows: the batch size of one training sample is 24, and the data set is iterated for 80,000 times in total. The stochastic gradient descent method is used as the network optimizer. The learning rate is initialized to 1E-3 and set to 1E-4 and 1E-5 respectively when the iteration reaches 60,000 and 70,000 times. Momentum is set to 0.9. In order to make the training process of the model more stable, the training of the model adopts the strategy of warm up learning rate automatic adjustment [24]. At the same time, the idea of transfer learning was introduced to load the weight parameters before the Conv4_3 layer which

had been trained in the VGG-16 network to accelerate the training of the network. The curve of the loss function is shown in Figure 10.

4.3. Analysis of Experimental Results. In target detection, mAP index is usually used to evaluate the comprehensive detection performance of a model, and the calculation of map value is related to the special-recall carve for each detection category. The abscissa and ordinate of the curve represent the recall rate and accuracy, respectively, and the area enclosed below is the average precision (AP). mAP is calculated by adding the AP values of each category and averaging them. Different from the classification task, the IOU values of the bounding box and the real box are used as the boundary to define the positive and negative samples in target detection. Values greater than the IOU are called positive samples, values less than the IOU are called negative samples, and the IOU value is usually set to 0.5. The test results of the test set in the rare-earth magnetic patch data set and the performance parameters before and after improvement are shown in Table 1.

According to the test results, the improved SSD algorithm in this paper has a 3.41% improvement in mAP compared with the original version of the SSD algorithm. This is mainly because the algorithm in this paper improves the ability of defect feature extraction and adopts the strategy of feature fusion between layers. The full use of contextual semantic information makes the rate of missed detection lower for small defects such as unfilled corners, which makes up for the deficiency of SSD algorithm in small target recognition ability to a certain extent. In addition to the enhancement of the backbone network embedded receptive field module, this algorithm also introduces the ECA network module to the low-resolution effective feature map of the top layer to obtain more effective features, which makes the detection rate of nonobvious defect targets such as line marks higher. The experiment found that using Mobile-NetV2 [20] as the backbone feature extraction network to detect the data of rare-earth magnetic patches can make the network have a high degree of lightweight and a fast
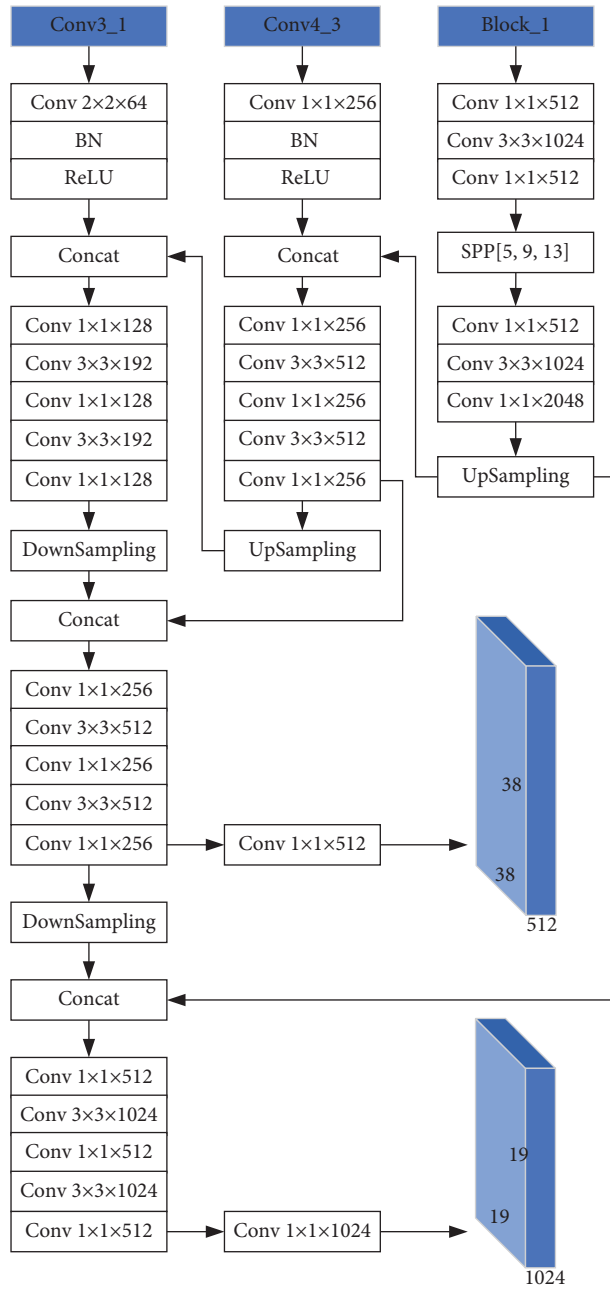
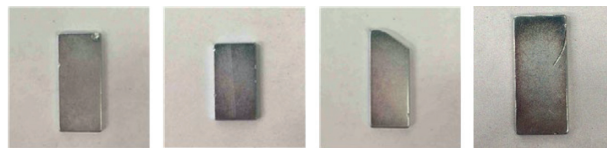FIGURE 8: Structure of feature fusion network based on bidirectional FPN.

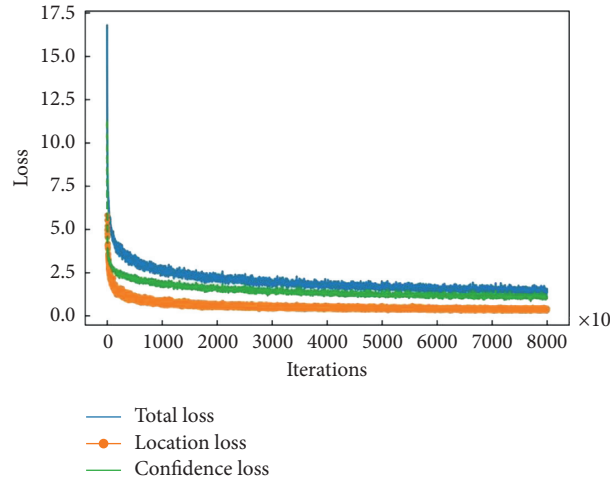FIGURE 9: Common defect appearance of rare-earth magnetic patch.

Figure 10: Loss function curve.

Table 1: Comparison of performance indexes between the improved SSD algorithm and other algorithms.

Algorithm	mAP/%	Detection accuracy (AP)/%			
		Unfilled corner	Crack	Deformation	Line mark
Faster R-CNN	77.48	69.86	85.11	63.51	91.45
YOLO-V3	81.88	86.58	84.63	66.25	90.06
MobileNetV2-SSD	73.83	62.66	79.74	67.50	85.43
SSD (original)	80.24	80.28	82.42	70.57	87.69
SSD (improvement)	**83.65**	86.37	85.31	70.89	92.02

detection speed, making it easy to deploy the device, but its detection accuracy is not reliable. Compared with the Faster R-CNN two-stage model that only uses the top layer to learn features, the improved algorithm in this paper has greater advantages. At the same time, compared with the more advanced one-stage model YOLO-V3 [24], defect recognition accuracy is slightly lower, but the comprehensive detection effect is better.

We perform ablation study to explore the effects of PANET bidirectional feature pyramid structure, parallel Multiscale Convolutional Layer, and efficient channel attention mechanism on detection accuracy. Here we investigate four models SSD, SSD + multiscale receptive field layer, SSD + multiscale receptive field layer + PANet, and SSD + multiscale receptive field layer + PANet + ECA. It can be seen from Table 2. After the enhancement of the backbone network embedded receptive field module and integration of the bidirectional feature pyramid in PANET for interlayer feature fusion, the accuracy of mAP was significantly improved to 81.37% and 83.24%. When combining the ECA module, we observed our best performance (83.65%).

In order to more intuitively evaluate the effect of SSD algorithm on defect detection before and after improvement, some experimental results are shown in Figures 11 and 12.

As shown in Figure 11, in addition to the improved SSD algorithm in this paper, the SSD algorithm with the lightweight MobileNetV2 convolutional neural network as the backbone feature extraction network and the original SSD algorithm can both locate and identify defects. The improved

SSD algorithm in this paper is more accurate in the bounding box (the detection results of line marks and deformation are shown in the figure).

In actual manufacturing, unfilled corner type defects are very common. The size of such defects is different, and small-sized unfilled corner defects often appear. The improved SSD defect detection algorithm shows stronger recognition ability for this type of defect, the algorithm has a lower rate of missed detection, and the recognition rate is significantly improved. At the same time, its recognition ability is also improved for defects with larger size but nonobvious features, as shown in Figure 12.

This paper compares the detection speed of SSD before and after improvement, which is 108 frames per second and 55 frames per second, respectively, on GeForceRTX2080TI high-performance graphics card, as shown in Table 3. Because the original SSD algorithm integrates the bidirectional feature pyramid structure and channel attention mechanism and enhances the trunk extraction network, the number of parameters is more than the original algorithm. Compared with the original algorithm, the proposed algorithm integrates the bidirectional feature pyramid in PANET for interlayer feature fusion. The algorithm has a large amount of computation, which improves the overall mAP and increases the inference time of the model. In the application of industrial video analysis and processing, the speed of reading image data from the camera is about 25 frames per second. The improved SSD target detection algorithm in this paper meets the requirements of real-time online detection while maintaining good detection performance.

TABLE 2: Ablation study: mAPs of different modules.

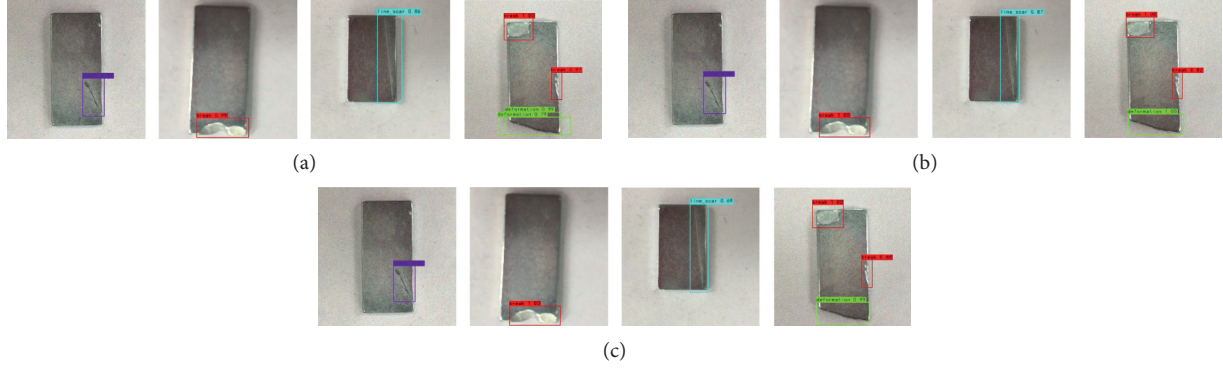| Structure	mAP/%
SSD	80.24
SSD + multiscale receptive field layer	81.37
SSD + multiscale receptive field layer + PANet	83.24
SSD + multiscale receptive field layer + PANet + ECA	83.65

(a)

(b)

(c)

FIGURE 11: The detection effect of SSD algorithm on large-size defects before and after improvement. (a) Detection effect of MobileNetV2-SSD. (b) Detection effect of original SSD. (c) Detection effect of improved SSD.
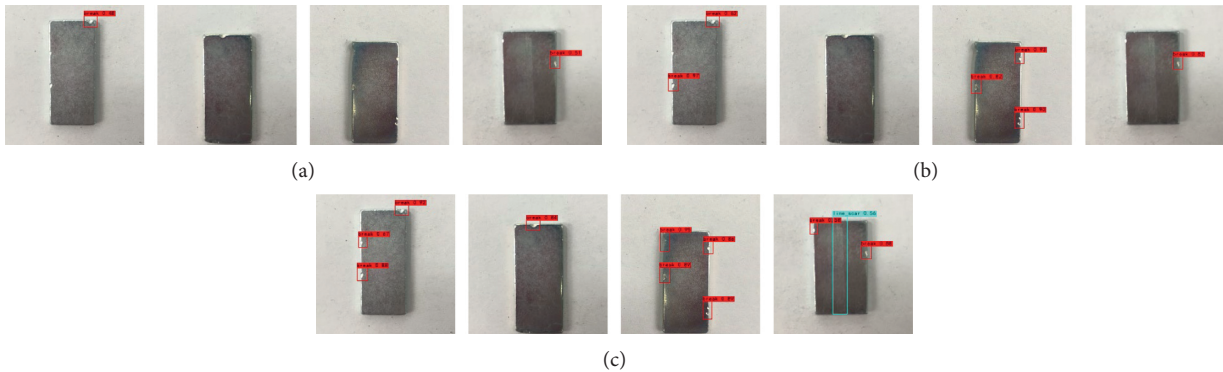

(a)

(b)

(c)

FIGURE 12: The detection effect of SSD algorithm on unfilled corners and line marks before and after improvement. (a) The detection effect of MobileNetV2-SSD. (b) The detection effect of original SSD. (c) The detection effect of improved SSD.

TABLE 3: Parameter comparison before and after improvement of SSD algorithm.

Algorithm	Parameter (M)	Computation (GMac)	mAP (%)	Speed (FPS)
SSD (original)	24.15	30.64	80.24	108
SSD (improvement)	71.09	57.63	83.65	55

5. Conclusion

This paper presented a defect detection method of rare-earth magnetic patch based on improved SSD. The SSD algorithm with a relatively balanced detection accuracy and inference speed was selected to detect magnetic patch defects and analyze the detection results. It was observed that SSD had a good identification effect for obvious defects with large size, but for defects with obscure features and small size, there was often a situation of missing inspection.

In order to further improve the detection accuracy of SSD algorithm, this paper embedded multiscale receptive field module into the backbone network of SSD algorithm to improve the feature extraction ability of the model, using a bidirectional feature pyramid idea to integrate high-level features and low-level features, combined with efficient channel attention mechanism to enhance the detection ability of the network. Experiments have proved that, compared with the original algorithm, the improved algorithm in this paper has a significant improvement in the

10

Complexity

recognition ability of small-size defects, with mAP reaching 83.65%. And the recognition ability of large-size defects with nonobvious features has also been enhanced. The detection speed can reach 55FPS on the experimental platform, which meets the requirements of online automatic detection.

Data Availability

The image data sets used in this paper were all collected from Guangxi Jinyuan Rare Earth Co., Ltd. (Guangxi Nonferrous Metals Group). The data sets were manually classified with the assistance of sorting workers and photographed on site with industrial cameras.

Conflicts of Interest

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

Acknowledgments

References

[1] N. Neogi, D. K. Mohanta, and P. K. Dutta, "Review of vision-based steel surface inspection systems," *EURASIP Journal on Image and Video Processing*, vol. 2014, no. 1, p. 50, 2014.

[2] X. Tang and Y. Wang, "Visual inspection and classification algorithm of rail surface defect," *Computer Engineering*, vol. 39, no. 3, pp. 25–30, 2013.

[3] X. Fang and Z. Shi, "Wafer defect detection and classification algorithms based on convolutional neural network," *Computer Engineering*, vol. 44, no. 8, pp. 218–223, 2018.

[4] H. Ponce, C. Cevallos, R. Espinosa, and S. Gutiérrez, "Estimation of low nutrients in tomato crops throughthe analysis of leaf images using machine learning," *Journal of Artificial Intelligence and Technology*, vol. 1, pp. 131–137, 2021.

[5] D. Jiang, G. Hu, G. Qi, and N. Mazur, "A fully convolutional neural network-based regression approach for effective chemical composition analysis using near-infrared spectroscopy in cloud," *Journal of Artificial Intelligence and Technology*, vol. 1, no. 1, pp. 74–82, 2021.

[6] C. Chang, C. Chang, C. Li, and M. Jeng, "Learning vector quantization neural networks for LED wafer defect inspection," in *Proceedings of the Second International Conference on Innovative Computing, Informatio and Control (ICICIC 2007)*, Kumamoto, Japan, September 2007.

[7] M. Yazdchi, M. Yazdi, and A. G. Mahyari, "Steel surface defect detection using texture segmentation based on multifractal dimension," in *Proceedings of the 2009 International Conference on Digital Image Processing*, Bangkok, Thailand, March 2009.

[8] K. Yildiz, A. Buldu, and M. Demetgul, "A thermal-based defect classification method in textile fabrics with K-nearest neighbor algorithm," *Journal of Industrial Textiles*, vol. 45, no. 5, pp. 780–795, 2016.

[9] Y.-J. Jeon, D.-C. Choi, S. J. Lee, J. P. Yun, and S. W. Kim, "Defect detection for corner cracks in steel billets using a wavelet reconstruction method," *Journal of the Optical Society of America A*, vol. 31, no. 2, pp. 227–237, 2014.

[10] P. Sun, Y. Zhao, and S. Zhu, "An approach to improve SSD through mask prediction of multi-scale feature maps," *Pattern Analysis and Applications*, vol. 24, no. 9, 2021, republish.

[11] Y. Huang, C. Qiu, and K. Yuan, "Surface defect saliency of magnetic tile," *The Visual Computer*, vol. 36, no. 1, pp. 85–96, 2018.

[12] C. Liu, J. Zhang, and J. Lin, "Detection and identification of surface defects of magnetic tile based on neural network," *Surface Technology*, vol. 48, no. 8, pp. 330–339, 2019.

[13] Z. Wen and M. Zhou, "Recognition of blowholes and cracks on surface of magnetic tile based on deep learning," *Ordnance Material Science and Engineering*, vol. 43, no. 6, pp. 106–112, 2020.

[14] C. Y. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg, "DSSD: deconvolutional single shot detector," 2017, https://arxiv.org/abs/1701.06659.

[15] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, https://arxiv.org/abs/1409.1556.

[17] J. Redmon, S. Divvala, R. Girshick, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," 2016, https://arxiv.org/abs/1506.02640.

[18] T. Y. Lin, D. Piotr, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," 2017, https://arxiv.org/abs/1612.03144.

[19] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," 2018, https://arxiv.org/abs/1803.01534.

[20] Y. Li, H. Huang, Q. Xie, L. Yao, and Q. Chen, "Research on a surface defect detection algorithm based on mobileNet-SSD," *Applied Sciences*, vol. 8, no. 9, p. 1678, 2018.

[21] S. Liu, D. Huang, and Y. Wang, "Receptive field block net for accurate and fast object detection," *Computer Vision—ECCV 2018*, vol. 11215, pp. 404–419, 2018.

[22] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-net: efficient channel attention for deep convolutional neural networks," 2020, https://arxiv.org/abs/1910.03151.

[23] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," 2017, https://arxiv.org/abs/1709.01507.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2016, https://arxiv.org/abs/1512.03385.