EasyChair Preprint

№ 4681

Surface Defect Detection Method of Hot-Rolled Steel Strip Based on Improved SSD Model

Xiaoyue Liu and Jie Gao

December 1, 2020

Surface defect detection method of hot-rolled steel strip based on improved SSD model

Liu Xiao-yue, GAO Jie

（c, College of Electrical Engineering, Tangshan 063210, China）

[Abstract] In order to reduce the influence of surface defects of hot rolled steel strip on product performance and appearance, a surface defect detection method combining attention mechanism and multi-feature fusion network is proposed. This method takes the traditional SSD network as the basic framework, selects RESnet-50 as the feature extraction network, and integrates the low-level features and the high-level features to complement each other, so as to improve the accuracy of detection. In addition, the channel attention mechanism is introduced to filter and retain important information, which reduces the network computation and improves the network detection speed. The experimental results on NEU-DET data set show that the accuracy of this method for surface defect detection of hot-rolled steel strip is obviously higher than that of traditional SSD network, and it can meet the real-time requirements of industrial detection.

[Key words] hot rolled strip; surface defect; SSD network; ResNet-50; feature fusion; Channel attention mechanism

Chinese Library Classification Number: TP391.41; **Document Marking Code:** A

Preface

Hot-rolled steel strip is an important material for industrial production and is widely used in machinery manufacturing, automobile production, aerospace and other fields. The surface quality has an important influence on the aesthetics, performance and durability of the product [1]. However, due to the poor actual production environment and the complexity of the process flow, hot-rolled steel strips are susceptible to many factors such as rolling equipment, processing technology, raw materials, and external environment during the production process, resulting in the formation of various types on the surface. Defects [2]. Because of this, the rapid and accurate detection of hot-rolled steel strips has become a key issue for many scientific research institutions at home and abroad.

Most of the current hot-rolled steel strip surface defect detection tasks are realized in two ways. The first is that the inspector detects the defects with the aid of the computer, but this method relies on the inspector's subjective judgment, and in the batch inspection process, the human eyes are prone to fatigue, leading to missed inspections and false inspections. It happened. The second is a detection method based on traditional machine vision. This type of method mostly uses image segmentation combined with artificially designed features and classifiers. However, this type of algorithm is difficult to achieve effective detection results in real and complex industrial environments, and generalization Poor ability [3].

In recent years, with the continuous development and progress of deep learning, many target detection algorithms have emerged. On the whole, the target detection algorithm based on deep learning is divided into two categories [4]: one-stage network and two-stage network. The core idea of the two-stage network represented by RCNN, Fast-RCNN, and Faster-RCNN is a method based on candidate regions. The candidate frame that may have a target is first generated, and then the target location and classification are further performed, and the detection accuracy is high. But the efficiency is low; while the one-stage network represented by YOLO and SSD

Received dat:： **Revision date:**

Introduction to the first author: Liu Xiaoyue (1965-), female, Han nationality, from Tangshan, Hebei, PhD, professor. Research direction: detection and control technology and intelligent devices. E-mail: 11250608@qq.com

Introduction to the second author: Gao Jie(1996-), male, Han nationality, from Tangshan, Hebei, PhD, professor. Research direction: detection and control technology and intelligent devices. E-mail: 330218779@.com

Fund Project: National Natural Science Foundation of China (51574102)

directly divides the area on the input image and performs the target detection task, taking into account the detection efficiency and detection accuracy. These algorithms have been improved by researchers and have been applied to detection work in various fields. The AFE-SSD model constructed by Jiang Jun and others based on SSD combined with the hole convolution and feature enhancement algorithm has significantly improved the detection accuracy of small targets [5]. Zhu Deli et al. used MobileNet as the feature extractor of VGG16-SSD, which improved the feature extraction capability and the robustness of the model [6]. Wu Shoupeng and others used the bidirectional feature pyramid to improve the Faster-RCNN network and improve the detection ability of multi-scale defects [7]. Although the above methods effectively improve the detection accuracy, they also add parameters to the network, making it difficult for the detection speed to meet the real-time detection requirements.

Considering the detection accuracy and speed of hot-rolled steel strip surface defects, this paper uses ResNet-50 instead of VGG-16 as the feature extraction network based on the SSD network, and merges the shallow features with the deep features , To make up for the shortcoming of insufficient shallow feature semantic information, and then filter the obtained feature images through the channel attention mechanism, so as to realize the effective use of important information and reduce the amount of calculation. Experimental results show that the detection accuracy of this method for small targets is improved, and the detection speed is significantly higher than that of SSD networks.

1 Typical surface defects of hot rolled steel strip

Due to the external environment, production and processing technology and the selection of raw materials, the surface of the hot rolled steel strip is prone to some defects. However, because the hot-rolled steel strip is an intermediate product, there is still further deep processing in the follow-up. Therefore, in the surface quality inspection of the hot-rolled steel strip, some defect types are not counted, so here is only for the six common hot-rolled steel strips The types of surface defects are introduced [8].

1) Rolled-in scale: During the rolling process, the iron oxide scale is pressed into the surface of the steel plate, generally in the form of strips, lumps or scales, and the color is brown or black.

2) Crazing: It is a relatively serious surface defect. Cracks of different shapes, depths and sizes will form on the surface of the steel strip, which will cause serious damage to the mechanical properties of the steel plate.

3) Inclusion: Divided into metallic inclusions, non-metallic inclusions and mixed inclusions. The surface of the steel strip presents a brown-red, yellow-brown or black embedded structure, and the inclusions are randomly distributed and different in shape.

4) Pitted surface: The partial or sheet rough surface formed on the surface of the steel plate will have small pits of different depths and shapes.

5) Scratches: It mostly occurs during the conveying process of rolled steel strips. The scratches appear brown or light blue at high temperatures, have irregular shapes, and generally have long lengths.

6) Patches: Appears as densely distributed approximately circular bright spots.

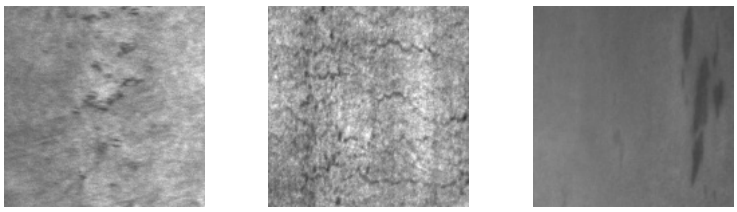
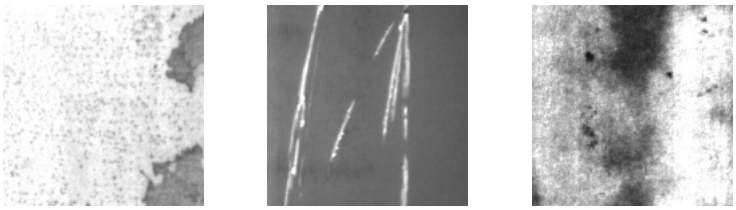


Fig.1 Surface defect diagram of hot rolled steel strip



pitted surface scratches patches

Fig.1 Continued

2 SSD model and improvement method

2.1 SSD model

The SSD target detection algorithm [9] is a multi-class single-order target detection algorithm proposed on the basis of the anchor generation method of Faster-RCNN [10] and the meshing idea of YOLO [11]. While maintaining the detection speed of the one-stage algorithm, this algorithm is also equivalent to the two-stage deep learning target detection algorithm in terms of detection accuracy, and is one of the current mainstream deep learning target detection algorithms.

The structure of the SSD network is shown in Figure 1. It uses VGG-16 as the basic network to perform feature extraction on the input image and adds 4 additional cascaded convolutional layers. The input image will generate a series of feature maps through the network, and the ones that are mainly used for the final prediction are the feature maps obtained by the six convolution layers of Conv4_3, Conv7, Conv8_2, Conv9_2, Conv10_2, and Conv11_2. Finally, the detection module is used to perform classification and regression calculations on each feature map to obtain the type of detection target and the position of the prediction frame in the feature image. Finally, the results obtained by the detection module are integrated, and the final detection result is obtained through the method of non-maximum suppression.

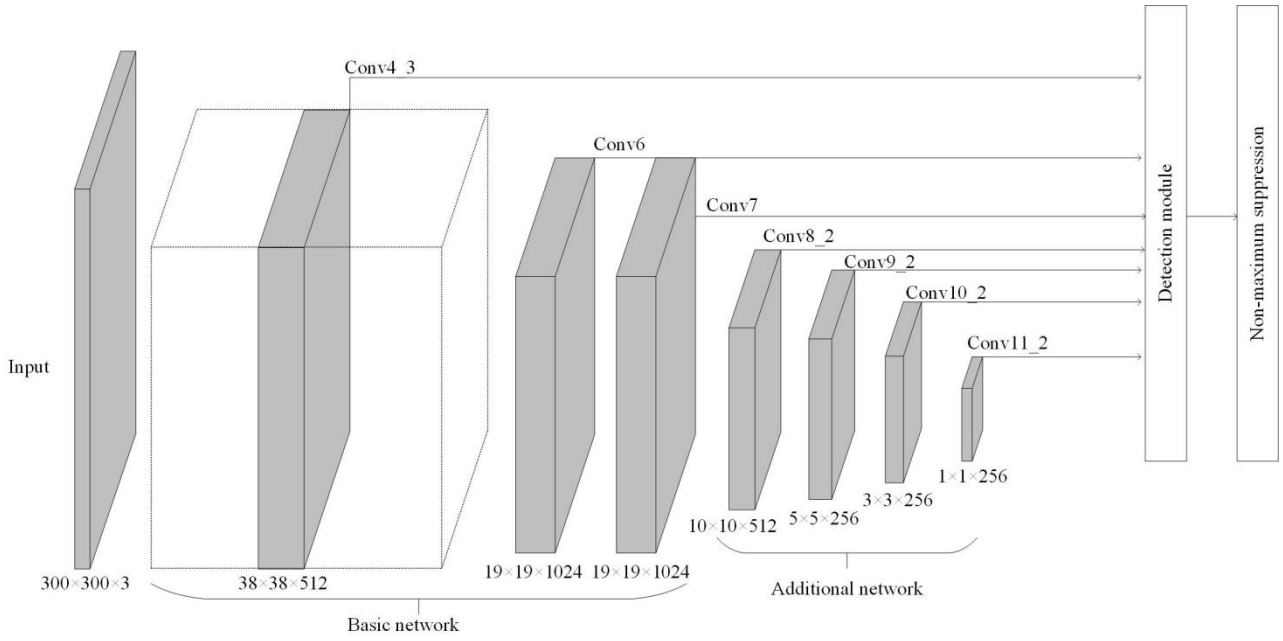


Fig.2 SSD algorithm structure diagram

1.2 Residual network

Although the deepening of the network can improve the feature extraction ability, the simple stacking of network layers will cause gradient explosion/disappearance and degradation. The degradation of the network is reflected in the decrease of training accuracy and test accuracy.

E et al. [12] proposed a residual network (ResNet) to solve the problem of network degradation. The principle of each residual learning module in the residual network is shown in Figure 2, X_1 Represents the input of the residual block, And the real output is $F(X_1)$, The expected output is X_{1+1}, Each module also superimposes the input on the output in a direct mapping manner, but the ReLU function needs to be used for activation first. After superposition, the output result becomes $F(X_1)+X_1$, The learning content of the network becomes $F(X_1)=X_{1+1}-X_1$ Residual form. If the number of layers of the network exceeds the optimal number of layers, the residual network will train the mapping of the excess layers as $F(X)=0$, That is, these layers become an identity map with equal input and output, so that network degradation can be avoided. Moreover, the residual network simplifies the learning goal, so that the network training can converge more quickly.

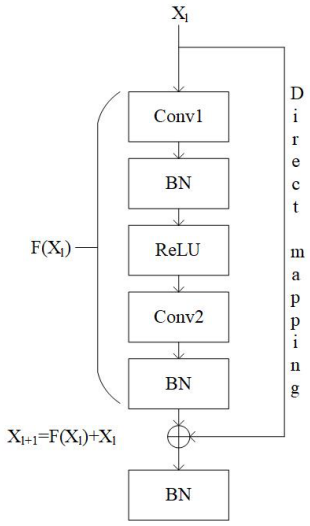


Fig.2 Residual learning module schematic diagram

Commonly used ResNet models include ResNet-50, ResNet-101, ResNet152, etc. Based on the consideration of model parameters, this paper uses ResNet-50 as the basic feature extraction network, and its network structure is shown in Table 1. The network is mainly composed of five residual learning modules. The 3×3 convolutional layer in the middle of each residual block first reduces the amount of calculation through a 1×1 convolutional layer, and then passes through another 1×1 convolution Layers are restored, which reduces the amount of calculation while ensuring accuracy.

Table 1 Resnet-50 network architecture

Layer	Structure	Output size
Conv1	7×7，64，stride 2	112×112
Pool1	3×3，max，stride 2	56×56

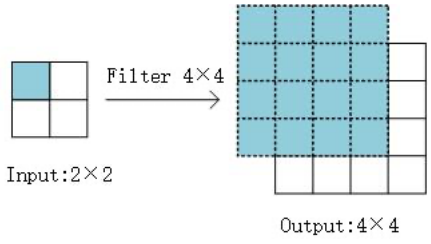
Conv2_x	$\begin{pmatrix}1\times1,64\\3\times3,64\\1\times1,128\end{pmatrix}\times3$	56×56
Conv3_x	$\begin{pmatrix}1\times1,128\\3\times3,128\\1\times1,512\end{pmatrix}\times4$	28×28
Conv4_x	$\begin{pmatrix}1\times1,256\\3\times3,256\\1\times1,1024\end{pmatrix}\times6$	14×14
Conv5_x	$\begin{pmatrix}1\times1,512\\3\times3,512\\1\times1,2048\end{pmatrix}\times3$	7×7
Pool5	Average pool，stride 1	2×2

1.3 Feature fusion

The SSD algorithm is a convolutional neural network based on forward propagation, which is hierarchical and can extract feature maps of different scales and different semantic information. In general, the feature map extracted by the shallow network has a higher resolution, but the receptive field corresponding to each feature point is small, and the semantic information is scarce. It is suitable for predicting small objects; the feature map extracted by the deep network has undergone many After the layer of convolutional pooling operation, the resolution is low, the receptive field corresponding to each feature point is larger, and the semantic information is rich, which is suitable for predicting large objects [13].

Most of the surface defects of hot-rolled steel strip are small-area defects, and the defective part occupies a low proportion of pixels, which is a small target detection. The detection of small objects requires that the features extracted by the network have higher resolution and richer semantic information. The SSD network uses multi-scale feature maps for target detection, which leads to its use in small object detection. The effect is not ideal. In view of this situation, consider the fusion of shallow feature maps and deep feature maps to improve the detection accuracy of the network for small targets.

However, the resolution of the feature maps extracted by different convolution layers may be different, so it is necessary to scale the convolution feature maps of different resolutions first, and then perform feature fusion after unifying the resolution. Upsampling is a method of converting low-resolution images into high-resolution images. In convolutional neural networks, the commonly used upsampling method is transposed convolution, also known as deconvolution, but it only realizes size adjustment, not value restoration in a mathematical sense. The principle of deconvolution is shown in Figure 4. Assuming that the input size is 2×2, a convolution kernel with a size of 4×4 is used in the deconvolution process, and the padding value is set to 1, as the convolution kernel takes 2 as the pace Move on the input image, you will get four output windows of 4×4 size, superimpose the overlapping parts of different output windows, and then remove the outermost padding value, the final output image size is 4×4, the size is enlarged 2 times the input.



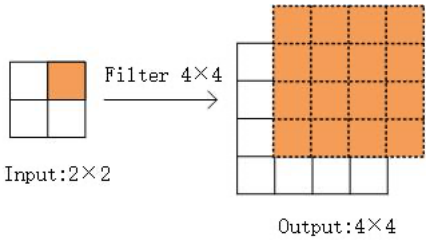


Fig.3 Deconvolution principle diagram

This article refers to the deconvolution module of the DSSD network [14]. Its structure is shown in Figure 3. First, the high-level and low-resolution images are up-sampled, and then the number of channels is unified through the 1×1 convolution layer, and then adjusted The later deep feature maps and shallow feature maps are fused, and finally the ReLU function is used for activation.

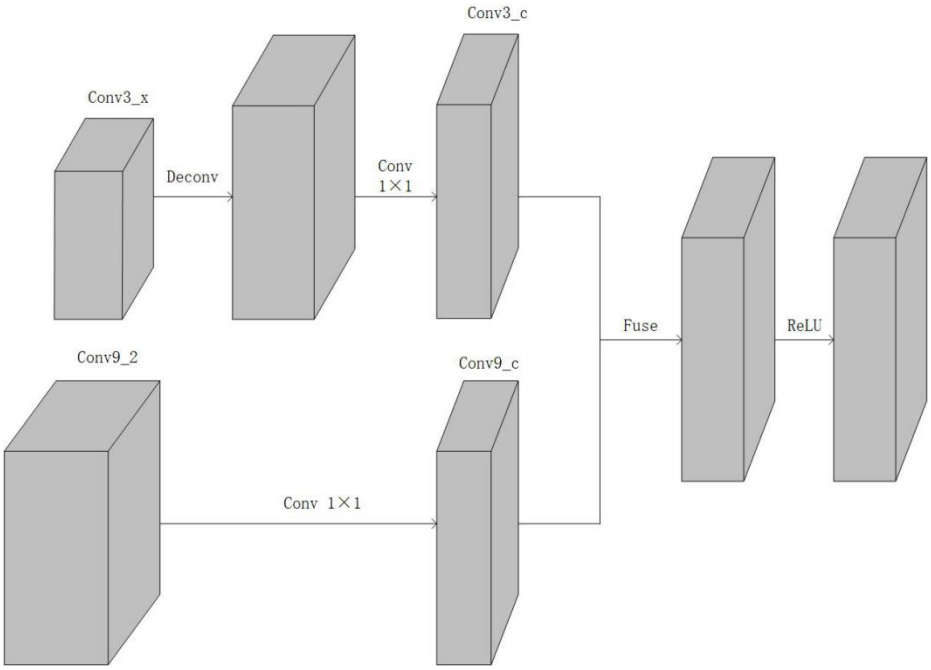


Fig.4 Feature fusion module

1.4 Attention mechanism

The working principle of the attention mechanism [15] is to establish a new layer of weights. After learning and training, the network learns the more important areas in each training image, and strengthens the weights of these areas to form The so-called attention. In this article, the channel attention mechanism module includes three parts, namely squeeze, encouragement and attention. The algorithm flow after adding the attention mechanism is shown in the figure:

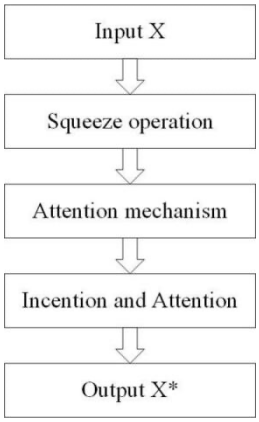


Fig.5 Flow chart of attention mechanism algorithm

The principle of extrusion is shown in formula (2):

$$S = \frac{\sum_{i=1}^{H}\sum_{j=1}^{W} X(\mathrm{i}, j)}{H * W} \tag{2}$$

Formula (2) is actually a global average pooling operation. In the formula, H and W represent the length and width of the input feature image X, (i, j) represents the point at position (i, j) on the image X, and C represents the number of channels in the image X, and each channel All the eigenvalues within are averaged and then summed to obtain a one-dimensional array S of length C. After the output S is obtained, the correlation between each channel is modeled, which is the process of excitation. The principle is shown in Equation 2:

$$R = S\mathrm{igmoid}(\mathrm{W}_2 \cdot \mathrm{ReLU}(\mathrm{W}_1\mathrm{S})) \tag{3}$$

The dimension of W_1 is C1×C1, the dimension of W_2 is C2×C2, where C2=C1/4, are these two weights trained through the ReLU function and Sigmoid function, and a one-dimensional excitation weight is obtained for each The layer channel is activated, and the obtained R has a dimension of C1×1×1. Finally, the attention calculation:

$$X^* = X \cdot R \tag{4}$$

Replace the original input X with the feature map X* obtained by the attention module, and send it to the improved algorithm model for detection. In other words, this process is actually a process of scaling, and different weights are multiplied by different channel values. Thereby enhancing attention to important channels.

2 RAF-SSD network

Through the improvement of SSD, the algorithm RAF-SSD in this paper is obtained. Its network structure is shown in Figure 5. First, ResNet-50 is used as the feature extractor, and then Conv3_x and Conv8_2, Conv7 and Conv10_2 are feature fused, and then combined with other The feature map is first filtered by the attention module for important information, then the detection task is performed, and finally the result is obtained through maximum value suppression.

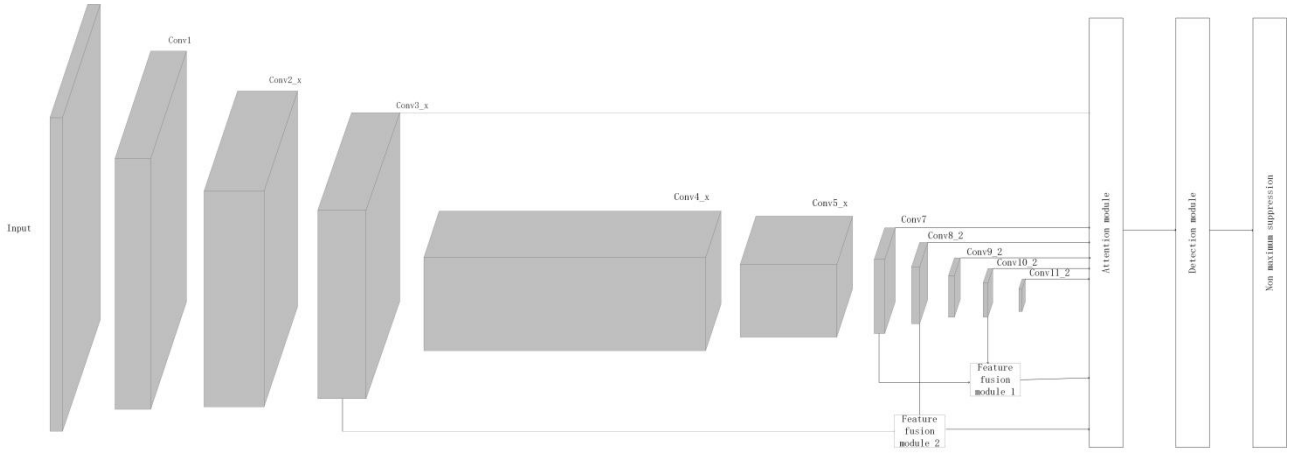


Fig.6 The improved network structure

3 Experimental data and results

3.1 Priori box setting and matching

The a priori box is set with scale and aspect ratio as two main aspects. The scale follows the principle of linear increase, and its size changes as shown in equation (5):

$$S_{\mathrm{p}}=S_{\min}+\frac{S_{\max}-S_{\min}}{m-1}(\mathrm{p}-1),\mathrm{p}\in[1,m] \tag{5}$$

Among them, S_p is the ratio of the prior frame to the image size; S_{max} and S_{min} are the maximum and minimum values of the ratio respectively; p is the current feature map; m is the number of feature maps.

Set S_{min} to 0.3 and S_{max} to 1.0 during training. The first feature map sets the a priori frame scale ratio as Smin/2, and the subsequent feature maps increase linearly according to the principle in equation (5). Set the aspect ratio of the training prior frame to {1,2,1/2,1/3}.

The matching principle adopted is: first, for each target frame in the picture, a prior frame with the largest cross and ratio is selected as the matching object, and this prior frame is identified as the prior frame of the positive sample. Among them, IoU is the intersection of a priori boxes. By setting the IoU value, the matching between the a priori box and the target box is realized.

3.2 Loss function

The loss function of the model is mainly divided into two parts: position loss and regression loss. The two parts jointly evaluate the detection effect of the network. The overall loss is shown in formula (6):

$$L(x,c,w,h)=\frac{1}{N}(\mathrm{L}_{\mathrm{conf}}(x,c))+\beta\, L_{loc}(x,w,h) \tag{6}$$

The first part of the formula $\frac{1}{N}(L_{\mathrm{conf}}(x,c))$ Represents a loss of location, Where N represents the number of matching boxes with the complete target box, and c represents the confidence level;Part two $\beta L_{loc}(x,w,h)$ Represents the regression loss, where x represents the center position of the target box, w represents the width of the box, and h represents the height of the box.

3.2 Experimental data and hyperparameters

The configuration used in this experiment is Win10 64-bit operating system, Intel(R)Core(TM) i7-10170U CPU and NVIDIA GeForce 1070 graphics card. Use the Python language to integrate the opencv library on the Tensorflow deep learning framework.

After building the model, first initialize the parameters, select small batch stochastic gradient descent (SGD) as the optimizer, set the learning rate to 0.0001, set the learning rate attenuation factor to 0.92, and set the number of training samples per batch (batch) It is 32, and the number of iterations of the training data set is set to 20000 times to obtain the final model. Then choose the mAP value as an index to measure the accuracy of the model.

The data used in this article is the open-source hot-rolled steel strip surface defect data set NEU-DET, which includes six types of hot-rolled steel strip surface defects, that is, scale pressure, cracks, surface inclusions, pits, scratches and Surface spots, in order to facilitate training and detection, use 0-5 corresponding to represent the above five types of defects. Each type of defect contains 300 samples, a total of 1800 sample pictures. In order to prevent the occurrence of over-fitting and improve the generalization ability of the trained model, the images of the data set are enlarged to 13,400 images through image enhancement methods such as flipping, random cropping, and adding noise. After that, the data set is divided into training data set and test data set according to the ratio of 4:1, which are used for model training and performance testing respectively.

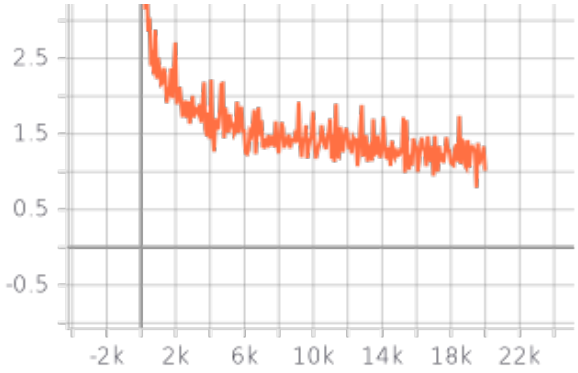


Fig.7 Curve of model training loss function

Figure 7 is the convergence curve of the loss value of the model during the training process. The horizontal axis represents the number of iterations of training the model using the training data set, and the vertical axis represents the total loss value. It can be seen that when the number of iterations reaches the 14000th time, the loss The size of the value remains around 1.3, and the change of the loss value tends to stabilize.

Then use the same data set to train the three models of YOLO-V3, SSD, and Faster-RCNN, and compare the performance of these three models on the test data set with the RAF-SSD model.

Table 2 Comparison of detection results of the three models in NEU-DET dataset

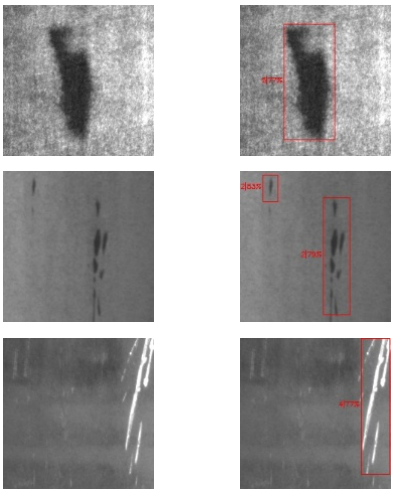
	YOLO-V3	SSD	Faster-RCNN	RAF-SSD
rolled-in scale	68.2	68.1	68.8	75.3
crazing	67.9	68.2	69.1	71.7
inclusion	71.5	72.4	74.3	75.5
pitted surface	68.4	68.4	69.2	72.6
scratches	68.3	67.8	68.1	75.4
patches	73.2	73.7	72.9	80.1
Average accuracy（mAP%）	69.6	69.8	70.4	75.1

It can be seen from the data in Table 2: The average accuracy of the algorithm in the six hot-rolled steel strip surface defects in the NEU-DET data set is 81.5%, which is compared to Faster-RCNN with the highest average accuracy in other models. The model is improved by 4.7%, which is 5.5% higher than the traditional SSD algorithm, which shows that the method in this paper can achieve effective detection of small targets and improve the accuracy of small target detection.

Table 3 Comparison of detection speed of the three models

Method	Detection Rate f/s^{-1}
SSD	54
YOLO-V3	52
Faster-RCNN	41
AF-SSD	53

The data in Table 3 shows that the detection speed of the RAF-SSD model is 53 frames per second, which is higher than the two-stage detection network Faster-RCNN, and compared to the two one-stage detection networks of SSD and YOLO-V3, the detection speed is basically the same . Considering that the feature fusion module is added to the model, the deconvolution operation and feature fusion are performed on the feature map extracted by the convolutional network. To a certain extent, the increased amount of calculation can indicate the attention added in the process of model improvement The model highlights important information, reduces update parameters, reduces the amount of parameters, and speeds up network detection.



（a）Original image　　（b）Detection result

Fig.6　Experimental detection effect drawing

From the inspection effect diagram, the defect area in the original image is marked by a rectangular frame, and the defect type and confidence level are also displayed. It shows that the improved algorithm proposed in this paper is accurate and effective in detecting surface defects of hot-rolled steel strips, and can detect surface defects of different types of hot-rolled steel strips.

4　Conclusion

In order to solve the problem of inaccurate positioning, poor robustness, and poor detection effect of the SSD algorithm in the small target detection task, the SSD network is selected as the basic framework, ResNet-50 is used as the feature extractor instead of VGG16, and the attention module is introduced. And the feature fusion module improves the algorithm and applies it to the surface defect detection of hot-rolled steel strip. The experimental results on the NEU-DET dataset show that the RAF-SSD model guarantees the detection speed. Compared with the traditional SSD algorithm, the detection effect of small target objects has been significantly improved, and it can meet the requirements of hot-rolled steel strip surface Testing the requirements for model accuracy and testing speed. However, the data set samples used in this article are small. The next step will be to enhance the data and expand the sample size, and then improve the performance of the network.

reference

1. Xu Ke, Wang Lei, Wang Jingyu. Surface defect recognition of hot-rolled steel plates based on tetrolet transform[J]. Instruments Science and Technology, 2016, 52(4): 13-19.
2. Wu Pingchuan, Lu Tongjun, Wang Yan. Nondestructive testing technique for strip surface defects and its applications[J]. Nondestructive Testing, 2000, 22(7): 312-315.
3. Wang Siyu, Gao Weixin, Zhang Xiangsong. Overview of Defect Detection algorithms in X-ray Images of Welding seams [J]. Thermal processing technology, 202,49(15): 1-8
4. Tao Xian, Hou Wei, Xu De. A survey of surface defect detection methods based on deep learning[J/OL]. Acta Automatica SinicamMonth,　2020-04-02.
5. Jiang Jun, Zhai Donghai. Single-stage object detection algorithm based on atrous convolution and feature enhancement[J]. Computer Engineering, 2020-06-18.
6. Zhu Deli, Lin Zhijian. Corn silk detection method based on MF-SSD convolutional neural network[J/ OL]. Journal of South China Agricultural University, 2020- 09-24.
7. Wu Shoupeng, Ding Enjie, Yu Xiao. Foreign body identification of belt based on improved FPN[J]. Safety in Coal Mines, 2019, 50(12): 127-130.
8. Yu He, Kechen Song, Qinggang Meng, Yunhui Yan. An End-to-end Steel Surface Defect Detection Approach via Fusing Multiple Hierarchical Features[J]. IEEE Transactions on Instrumentation and Measurement, 2019-05-08.
9. Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[C]// European Conference on Computer Vision. Springer International Publishing, 2016: 21-37.
10. Ren S Q, He K M, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal

networks[J]. In: Advances in Neural Information Processing Systems (NIPS). Montreal, Quebec, Canada: MIT Press, 2015, 91−99.
11. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016. 779−788
12. He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
13. Lei Pengcheng, Liu CongTang Jiangang, et al. Hierarchical feature fusion attention network for image super-resolution reconstruction, Journal of Image and Graphics, 2020(9): 1773-1786.
14. Fu C Y, Liu W, Ranga A, et al. DSSD: Deconvolutional single shot detector[J]. arXiv:1701.06659v1, 2017.
15. Hu J, Li S, Gang S. Squeeze-and-excitation networks[J]. arXiv preprintarXiv: 1709. 01507, 2017.