International Core Journal of Engineering Volume 8 Issue 1, 2022
ISSN: 2414-1895 DOI: 10.6919/ICJE.202201_8(1).0015

Application of Faster-Rcnn based on Resnet50 in Appearance Inspection of Industrial Products

Ao Duan[a], Yuefan Cong[b]

School of Electronic Engineering, Tianjin University of Technology and Education, Tianjin 300222, China

[a]duanao611@163.com, [b]826596556@qq.com

Abstract

Aiming at the problems of low recognition rate of surface defects of industrial products by traditional detection algorithms and inaccurate localization of small defects, an improved Faster RCNN deep learning network was proposed to detect defects. Firstly, after data enhancement, the traditional Faster RCNN feature extraction network is improved to make the network layer deeper to enhance the feature extraction capability of small defects. Then, the ROI Align algorithm is used to replace the rough ROI Pooling algorithm to obtain more accurate defect location information and obtain anchor frames that are more suitable for defects. Experimental results show that the recognition effect of the improved network on surface defect detection is up to 99%, which is more than 10% higher than the original Faster RCNN network, and the detection ability of small defects is also significantly improved.

Keywords

Deep Learning, Target Detection, Convolutional Neural Network, Faster RCNN, Resnet50.

1. Introduction

Machine vision in recognition, measurement, defect detection, positioning guide application has become more common, especially in defect detection scene, such as the lack of material, crack, material defects, such as in the case of accuracy is not high, machine vision, can completely replace human eyes and detecting accuracy can get very good repeatability [1]. Detection accuracy is very high, however, some defects and presents diversity and unpredictability, traditional machine vision algorithm have been unable to meet the application requirements, with the development of the depth of the artificial intelligence learning algorithm, and its defect detection application of machine vision, is to provide the appearance defect detection accuracy is one of the important methods of [2].

In 2004, Pernkopf[3], an American scholar, obtained the data range of steel ingot surface based on optical tomography, drew depth images using relevant 3D characteristics and sent them to The Bayesian network for feature extraction and classification, and the classification accuracy of segmented surface exceeded 98%. Yang Lijian et al. [4] from Shenyang University of Technology proposed a steel plate crack detection method based on balanced electromagnetic technology. It is verified that under the condition of AC excitation, the U-shaped sensor can effectively detect the transverse and longitudinal cracks on the steel plate surface by using the balanced electromagnetic technology, and can effectively distinguish the steel defects with crack types. Zhou Peng et al. [5] proposed a metal surface defect recognition method based on the fusion of shear wave and wavelet feature, respectively detected 6 kinds of defects of 3 kinds of metals and achieved good results, which verified that the method could detect different kinds of metal defects. Cheng Xingzhen [6] from Harbin Institute of Technology simulated and built a multi-mode non-destructive testing system for

ISSN: 2414-1895 DOI: 10.6919/ICJE.202201_8(1).0015

metal material defects based on the principle of non-destructive testing for defects. The system can detect surface and shallow surface cracks of aluminum and steel plates, and comprehensively and accurately detect the type, scale, depth and other information of metal defects. Feng Xuwen [7] from Qilu University of Technology designed a set of aluminum surface defect detection system, which combined median filtering to reduce noise and adaptive threshold to segment aluminum plate defects to preprocess defect images and extract feature data. The classifier adopted support vector Machine (SVM). This method can not only identify four common surface defects of aluminum plate with good results, but also identify new defect types.

With the booming development of deep learning, metal surface defect detection combined with deep learning has also been applied [8] .Such as: Wang et al. [9] designed a multi-layer convolutional neural network (CNN) to detect the defects of six metal defects, sampled the features of the original image through the sliding window method, and then dicclassified the small image blocks sampled from each type of image and compared them with the traditional method. The recognition effect is better than the traditional method. Mei et al. [10] proposed a convolutional denoising auto-encoders (CDAE) detection method for texture image defects combining the idea of image pyramid hierarchy and convolutional denoising auto-encoders. Compared with other traditional methods, the detection effect of this method is better on the atlas of cloth and silk fabric with strong repeatable background texture, but it is not ideal on the surface data set of metal surface machined parts.

In recent years, domestic scientific research institutions have carried out corresponding research on the application of machine vision to high-precision detection of structural parts' characteristics and defects. Researchers from Wuyi University developed a detection model based on machine vision for the diameter and width of the disk paper contour. The detection model is based on The Hough circle detection and Canny edge detection algorithm. The center coordinates, diameter and width of the disk paper contour are detected by extracting the disk paper contour. The contour skeleton extracted after graphic refinement is used to determine the position of feature points of edge pixels. The minimum quadratic function multiplication is used to fit the coordinates of feature points to calibrate the limit deviation and determine the conformity of products. The width detection error is about 0.8mm. The researchers from Southwest Jiaotong University studied the aircraft rivet size detection technology based on machine vision. OpenCV was used as the development platform, and the thresholding segmentation algorithm and watershed algorithm were used for image segmentation. At the same time, the image is processed by smoothing, morphology and contour extraction algorithm. The measured results show that the measurement accuracy can reach 0.01mm.

With the gradual rise of deep learning, technologies focusing on target detection and image recognition have become the research direction in which most researchers are interested [13]. With the progress of related technologies, the research on target detection is also deepening. Traditional target detection algorithms mostly use sliding window method or image segmentation to generate a large number of candidate frames, and then carry out feature extraction of candidate frames (including HoG[14], SIFT[15], Haar[16], etc.). Finally, the output features are transmitted to classifiers (SVM [17], Adaboost[18], Random[19],etc.) to judge the categories, and the accuracy and speed of detection are not high. Therefore, the traditional target detection algorithm is difficult to meet the requirements of commercialization. At the same time, with the development of society, products under the conditions of industrial production are more and more delicate, and the application scenarios of detection system are more and more abundant, such as uav aerial photography, street intersection monitoring, etc. In this condition, there are a lot of small targets in the image, so the ability of small target detection is particularly important.

ResNet network has been widely used in various computer vision tasks in recent years and has achieved outstanding performance. In this paper, RES-Net50 is selected as the feature extraction network of FtP-RCNN. By using residual network, input information is directly derouted to output to protect the integrity of information. The whole network only needs to learn the part of the difference

ISSN: 2414-1895 DOI: 10.6919/ICJE.202201_8(1).0015

between input and output to simplify the learning target and difficulty. And effectively solve the deep CNN model difficult training problem.

2. Faster RCNN Algorithm

Due to its high detection accuracy, the faster-RCNN algorithm has become one of the mainstream target detection algorithms. Compared with the YOLO series algorithms, the faster-RCNN algorithm is slightly insufficient in speed and has a high average detection accuracy (mAP). It partially integrates region proposal extraction and FAST-RCNN into a network model (RPN layer of region generation network), and its overall framework is shown in Figure 1:

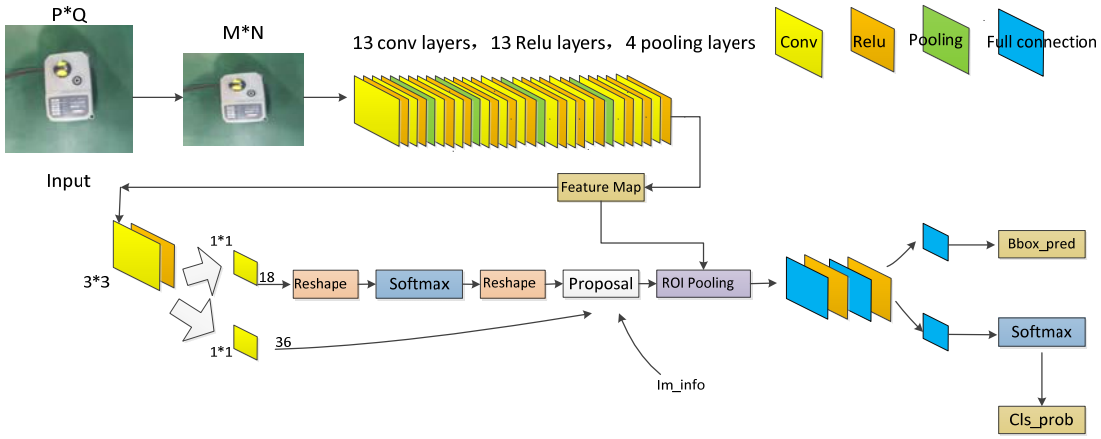


Figure 1. Internal structure diagram of Faster RCNN network

The algorithm can be roughly divided into feature extraction layer, region suggestion layer (RPN), ROI pooling layer, classification and regression. The detailed steps are as follows:

(1) Firstly, feature extraction network is used for feature extraction of input: feature extraction network is usually composed of convolution layer, pooling layer and activation layer.

(2) The generated feature map is transmitted to the RPN network to generate the suggestion box, and then the binary classification is performed to determine whether the target is included. Meanwhile, the feature map is transmitted to the ROI pooling layer for pooling operation, and the feature map of the candidate area with a fixed size is generated.

(3) Classify and regression the generated candidate region feature map to get the type and location of the object.

Compared with previous algorithms, telfa-RCNN has several major improvements: its core is the proposed RPN (Region Proposol Network) Network, which replaces the traditional method of generating candidate regions, realizes end-to-end training, and unify the whole object detection process into the same neural Network. RPN and Fast RCNN realize the shared convolution feature and reduce the training time. Use ROI Pooling to fix ROI on feature maps to feature maps of a specific size using maximum Pooling: Use NMS(non-maximum suppression) technology to screen the number of candidate boxes.

3. Improved Details of the Ftp-RCNN Algorithm

3.1 Feature Extraction Network

ResNet50 is used to replace the original VGG16 feature extraction network in feature extraction stage. As shown in Figure 2, RestNet50 contains 49 convolution layers and 1 full connection layer. Idblocks in the second to fifth stages are residual blocks without changing dimensions, and convblocks are residual blocks with added dimensions. Each residual block contains 3 convolution layers, so 1+3×(3+4+6+3)= 49 convolution layers. CONV is convolution operation, and Batch Norm is

ISSN: 2414-1895 DOI: 10.6919/ICJE.202201_8(1).0015

regularization processing, also known as BN layer, ReLU is activation function, MAXPOOL and AvgPOOL are maximum pooling layer and average pooling layer. Figure 3 shows that ResNet50 is stacked with multiple residual blocks, with which the deep network can be trained.

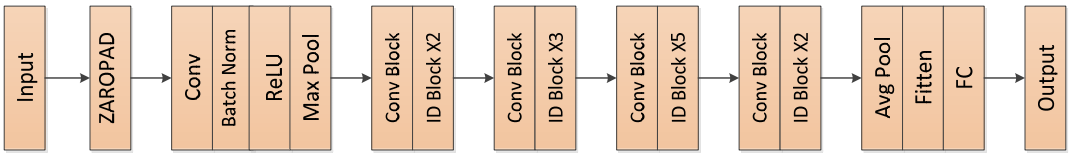


Figure 2. network structure of DensNet50

The residual block is shown in Figure 3:

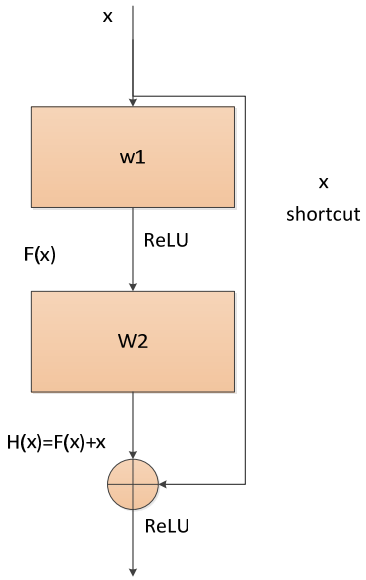


Figure 3. Fast residuals

When there is no shortcut (i.e. arrow on the right from X to ○), the residual block is an ordinary two-layer network, where the learning goal is $F(x) = H(x)$.The output of the second layer network before activating the function is $H(x)$. If the optimal output in this two-layer network is input x, then for a traditional network without shortcut, it needs to be optimized to $H(x) = x$; For the residual network structure, The learning objective is $F(x) = H(x) - x$. If the optimal output is x, we only need to optimize $F(x) = H(x) - x$ to 0. That is, the network only needs to make $H(x) = x$, that is, the identity mapping, which is defined in formula (3.1) and formula (3.2).

$$F(x) = W_1 g(W_2 x) \quad \text{formula(1)}$$

$$H(x) = F(x) + x \quad \text{formula(2)}$$

Where, W_1 and W_2 represent the weights of the first and second layers, and represent the ReLU activation function. If $F(x) = 0$, $H(x) = 0$ can be converted into the identity mapping function. At this point, the problem is transformed into a residual function problem, which is defined in Formula (3.3).

$$F(x) = H(x) - x \quad \text{formula(3)}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202201_8(1).0015

In ResNet50 residual network, a convolutional neural network with a depth of 50 layers is constructed by stacking 16 residual blocks, which makes the network have stronger feature extraction ability and higher recognition accuracy. In addition, the shortcut connection in residual block adopts identity mapping, so that the output of one layer can be used as the output of the following layer over several layers to directly carry out identity mapping operation without introducing additional parameters and computational complexity. Therefore, ResNet50 residual network containing 16 residual block stacking is used for pre-training.

3.2 Selection of Activation Function

When using Faster RCNN for ROI extraction, it was found through experimental tests that the model trained by the original method had missed detection and error detection in image rotation and small-scale ROI extraction. In this paper, FasterRCNN method is improved and optimized from activation function and network model respectivel.

The RELU function image is shown in Figure 4. Relu activation function can make network training faster: compared with the derivatives of Sigmoid and TANH, it is easier to calculate. Back propagation is the process of constantly updating parameters, because its derivatives are not complicated and its form is simple. RELU function increases the nonlinearity of the network. The function itself is a nonlinear function, which can be a grid fitting nonlinear mapping when added to the neural network. RELU activation function can converge quickly and has left soft saturation, which effectively alleviates the problem of gradient disappearance. When the value is too large or too small, the derivative of sigmoid and TANh is close to 0, while RELU is an unsaturated activation function without such phenomenon. Finally, the RELU function makes the grid sparsity and performs well in the absence of unsupervised pre-training, providing the sparse expression ability of the neural network. Therefore, RELU is selected as the activation function of the convolutional neural network.

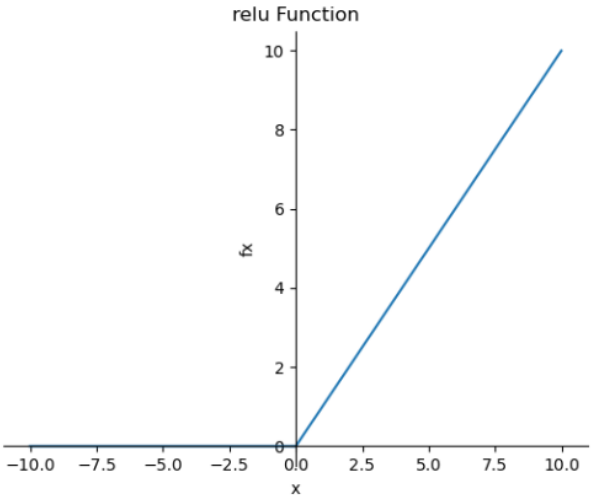


Figure 4. RELU function

3.3 Design of Loss Function

In order to achieve good results in images with large input aspect ratio differences, it is particularly important to balance the losses in classification and regression for this experiment. After the Rol Pooling layer, the output anchors are usually 256, and the resolution of feature map is the value generated after feature extraction network. In this piston surface defect detection, the input picture is 800×800. Therefore, after the down-sampling rate is 16, the size of the feature map obtained is (800/16)×(800/16)=2500. According to the loss function of the Faster R-CNN network, it can be known that the total number of samples, namely the number of Anchor, is 256, and the resolution of

ISSN: 2414-1895 DOI: 10.6919/ICJE.202201_8(1).0015

the feature map is, In order to achieve the balance loss condition, the value of the parameter is =10, so the loss function used this time is shown in Formula 3.4:

$$L(\{P_i\},\{t_i\}) = \frac{1}{N_{cls}}\sum_i L_{cls}(P_i, P_i^*) + \frac{10}{N_{reg}}\sum_i P_i^* L_{reg}(t_i, t_i^*) \quad \text{formula(4)}$$

Where, P_i and P_i^* are the truth value and forecast category of each Anchor category respectively, and $\sum_i L_{cls}(P_i, P_i^*)$ is the classification loss. $\sum_i P_i^* L_{reg}(t_i, t_i^*)$ is the regression loss, where t_i is the offset truth and t_i^* is the predicted offset.

4. Experimental Environment and Data

4.1 Design of Loss Function

4.1.1 Experimental Image Collection

The experimental images were captured by manual shooting, and the industrial products came from Tianjin Teck Actuator Co., LTD. Figure.5 shows the real scene of the company.

Figure 5. Production testing workshop diagram

4.1.2 Image Enhancement

To use deep learning technology to do detection tasks, it is necessary to have enough training samples to achieve a better recognition model. However, the images collected on site are very limited, so data enhancement of the original defect images is required before training , also known as image amplification. The network model trained by data enhancement has generalization energy also known as image amplification. The network model trained by data enhancement has more generalization ability. Basic operation methods include: left and right flip, up and down flip, rotation, zoom, translation, cropping, random occlusion, contrast enhancement, etc. Advanced operation methods include: blur, brightness adjustment, chroma adjustment, distortion, sharpening, noise disturbance, etc. There are many methods for data amplification, but it is not necessary to use all of them. Blind data amplification will bring useless data to the network, which is not beneficial to model training.

Choose the appropriate amplification operation according to the situation that may occur in the real scene. Because the valve is plastic products, in the light source is unstable and far from the detection conditions, easy to appear overexposure or dim, so you can choose brightness adjustment for amplification. From the shape of the consideration, the valve is not uniform hexahedron, because of the position of the uncertain, can choose up and down, left and right flip and rotation operation for amplification. Examples of amplification in this paper are shown in the figure below:

ISSN: 2414-1895 DOI: 10.6919/ICJE.202201_8(1).0015

(a) Original drawing

(b) 180 degree flip

(c) Left and right rotation

Figure 6. Contrast of image enhancement

4.1.3 Image Annotation

Different target detection models require different labels. Mainstream label formats include MAT, XML and TXT, etc. The contents encapsulated in label files of different formats are roughly the same

ISSN: 2414-1895 DOI: 10.6919/ICJE.202201_8(1).0015

and can be converted to each other without affecting the training effect. The image of this experiment was annotated by the image data annotation platform of Tianjin Huada Science and Technology co., LTD. The labeled image is shown in Figure 7, which is the defect map needed for detection in this paper.

The images collected by visual detection equipment can be obtained. There are three kinds of appearance defects to be detected, which are "scratch", "label" and "screw" respectively. In the image labeling and code, "scratch", "label" and "screw" are used respectively.

Figure 7. Label annotation

Figure 8. Marking scratches

Figure 9. Screw labeling

ISSN: 2414-1895 DOI: 10.6919/ICJE.202201_8(1).0015

The obtained annotation file contains five groups of key parameters, namely, the normalized value of the x coordinate of the label center point, the normalized value of the Y coordinate of the label center point, the normalized value of the width of the region of interest (ROI) W, the normalized value of the height of the region of interest (ROI) H, and the labeling parameters of the category. The key information in the document obtained after specific annotation is shown in Figure 10.

Figure 10. Key information in the data annotation file

4.1.4 Construction of Data Set

200 valid images were selected as the test set, and 800 images after image enhancement were selected as the training set.

4.2 Experimental Environment

The hardware configuration of the experiment was artificial intelligence all-in-one machine of Tianjin Huada Science and Technology. The processor was Ubuntu16 system, cpu-I5, 16GB memory, 250GB solid-state memory, NVIDIA RTX 2070S, python3.7 language environment, and the development tool was PyCharm. The development framework is Pytorch, the version of OpenCV is CUDA9.0, and the GPU acceleration library is CUDNN7.0.5.

4.3 Experimental Results and Analysis

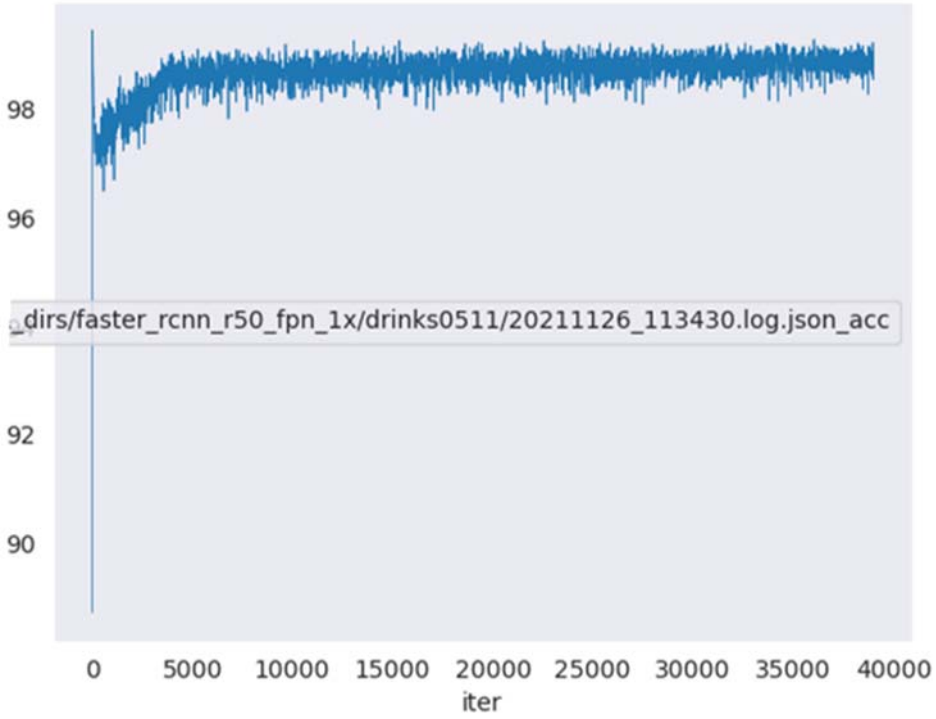
In order to better evaluate the performance of the network model, the model parameters were firstly selected, the training iteration times were set as 40000, and the learning rate was set as 0.0005. In the RPN network, enough proposals generated can avoid missing detection of defects to a certain extent, but all proposals used for subsequent training will reduce the training speed of the network and increase the training calculation burden. Therefore, the non-maximum suppression algorithm NMS is used to complete the selection of the proposal. Here, the non-maximum suppression threshold parameter in RPN network training is set to 0.7, and the number of proposals after NMS is set to 1000. See Table 1 for other detailed parameters.

Table 1. Network parameter configuration

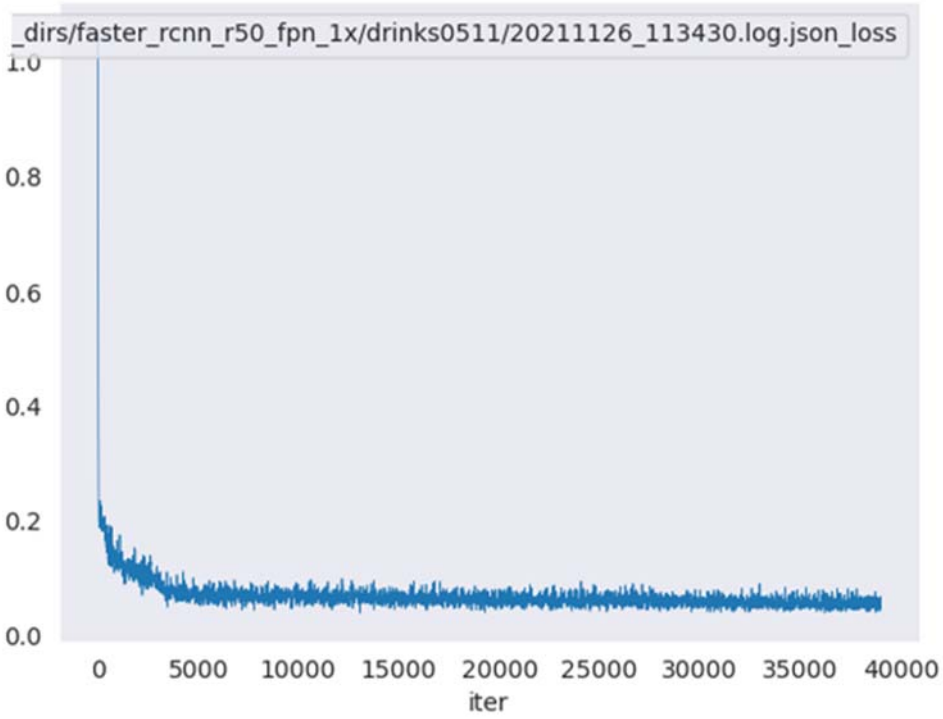
Backbone	Depth	Stage	Image_resize	Learn rate
Resnet50	50	4	300x300	0.0005
momentum	Rpn batch_size		Rpn_proposal_train	Rpn_proposal_test
0.9	256		1000	500

ISSN: 2414-1895 DOI: 10.6919/ICJE.202201_8(1).0015

The improved Faster RCNN recorded a loss function value every 50 iterations and trained the data using the prepared data set model. The final accuracy rate and total loss are shown in Figure 11.



(a) Accuracy

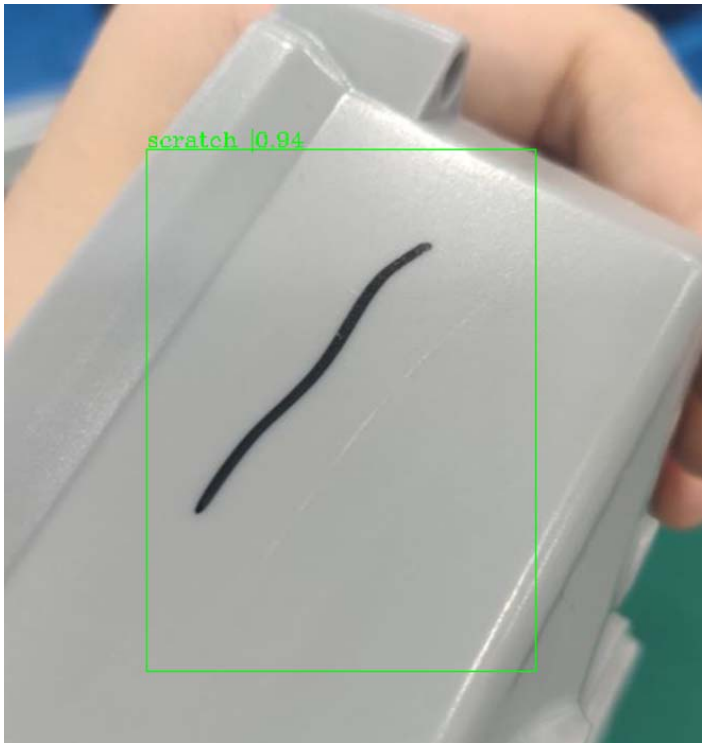


(b) Loss rate

Figure 11. Result diagram of improved model

The recognition effect is as follows:

ISSN: 2414-1895 DOI: 10.6919/ICJE.202201_8(1).0015



(a) The detection effect of scratches

(b) The detection effect of labels

Figure 12. Detection effect

5. Conclusion

In this paper, the deep learning network of improved Faster RCNN is used to identify product surface defects, and the following two improvements are made to the Faster RCNN network:

(1) The feature extraction network of the Faster RCNN network was changed from VGG to Resnet50 network, and the recognition ability of small surface defects was increased by introducing residual network. And the accuracy can be improved by adding considerable depth.

(2) Replace ROI Pooling by ROI Align to eliminate the quantization error generated by ROI Ooling, provide more accurate candidate boxes and obtain higher accuracy.

Finally, through the experimental comparison with the original Faster RCNN algorithm in the test set, it is verified that the improved Faster RCNN network has increased the recognition accuracy of

ISSN: 2414-1895 DOI: 10.6919/ICJE.202201_8(1).0015

different detection targets by 0.1-0.6 under the condition of little difference in detection speed, and has better defect detection effect than the original Faster RCNN. It provides an effective application reference value for industrial product appearance inspection.

References

[1] LIU J AN, Ll D J,SHAN H ZH.The current situation,technological innovation anddevelopment trendofaluminum-magnesium alloy extrusion pcess equipmentin my country [J]. Aluminum Processing, 2015 (1) :51-54.

[2] ZHANG H L. Research on key technologies of on-ineinspection system for aluminum plate surface defects [D]. Jinan: Qilu University of Technology, 2014.

[3] PERNKOPF F. Detection of surface defects on raw steel blocks using Bayesian network classifiers [J]. PatternAnalysis& Applications,2004, 7(3) : 333-342.

[4] YANG LJ, ZHENG W x,GAo s w , et al. Steel platecrackdefect detection method based on balanced electromagnetic technology [J]. Chinese Journal of Scientific Instrument, 2020 ,41(10): 196-203.

[5] ZHOU P, XU K, LIU SH H. Metal surface defectrecognition method based on the feature fusion of shearwave and wavelet [J]. Chinese Journal of MechanicalEngineering, 2015 ,51(6) : 98-103.

[6] CHENG X ZH. Research on non-destructive testing methodsfor metal material defects based on multi-modal signals [D]. Harbin: Harbin Institute of Technology, 2016.

[7] FENG x w. Research onthe recognition andclassification of complex defects on the surface ofaluminum plates [D]. Jinan: Qi lu University of Technology, 2018.

[8] LAN JH,WANG D,SHEN X P. Research progress ofconvolutional neural network in visual imagedetection[J]. Chinese Journal of Scientific Instrument, 2020, 41(4): 167-182.

[9] WANG T, CHEN Y , QIAO M, et al. A fast and robust convolutional neural network-based defect detection modelin product quality control [J]. International Journal ofAdvanced Manufacturing Technology, 2017, 94 (9) :3465-3471.

[10] MEI S, YANG H, YIN Z. An unsupervised-leaming-based approach for automated defect inspection ontextured surfaces[J]. IEEE Transactions onInstrumentation and Measurement,2018: 1266-1277.

[11] ZHONG J J, SHAO H,NIE ZH Y, et al. GPU-accelerated path-dependent digit al imagecorrelationmethod [J]. Joumal of Electronic Measurement andInstrument,2020,34 (10): 17-24.

[12] CAI CH P. Research on detection and classification ofmetal shaft surface defects based on deep learning [D]. Hangzhou: Zhejiang University of Technology, 2019.

[13] Wang P, Liu R, Xin X J, Liu PD. Multiscale Feature Fusion-Based Object Detection Algorithm[J/OL]. Laser & Optoclectronics Progress,2021, 58(2): 0210001.

[14] Dalal N, Triggs B. Histograms of oriented gradients for human detection[A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition[c]. USA: IEEE,2005.886-893.

[15] Lowe D G. Distinctive image features from scale-invariant key points [J]. International Journal of Computer Vision, 2004, 60(2) : 91-110.

[16] Lienhart R, Maydt J. An extended set of haar-like features for rapid object detection [A]. Proceedings of the International Conference on Image Processing [c].USA: IEEE,2002. 900-903.

[17] Lienhart R, Maydt J. An Introduction to Support Vector Machines and Other Kemel-based Learning Methods [M] .England: Cambridge University Press,2000.

[18] Freund Y, Schapirer E. Experiments with a new boosting algorithm[A]. Intemational Conference on Machine Learning [C]. USA: IMLS,1996.148-156.

[19] Liaw A, Wiener M. Classification and regression by random-forest [J] . RNews,2002, 2(3) : 18-22.