International Core Journal of Engineering Volume 7 Issue 6, 2021
ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0048

Application of Faster R-CNN Algorithm in Weld Position Recognition

Xu Wang[1], Houjun Lu[2]

[1]School of Logistics Science and Engineering, Shanghai Maritime University, Shanghai 201306, China;

[2]School of Logistics Engineering, Shanghai Maritime University, Shanghai 201306, China.

Abstract

The position recognition of welds in complex industrial environments is limited by traditional recognition methods and is susceptible to interference from the external environment. The Faster R-CNN algorithm model has a high recognition rate, but it has a large amount of calculation and a slow rate. Therefore, an improved algorithm based on the Faster R-CNN model is proposed. First introduce the composition of the industrial weld position recognition system and the deep learning neural network model, then explain the improved Faster R-CNN algorithm, introduce the number of regions adjustment layer and residual network, and finally conduct data experiments on the HF data set and working platform. The research results show that the proposed improved algorithm has the advantages of high recognition rate, fast speed, small model and diversified sample recognition, and can be deployed in the recognition of weld positions in industrial welding sites.

Keywords

Deep Learing; Target Recognition; Residual Network; Area Number Adjustment Layer; Industrial Weld.

1. Introduction

In the field of industrial welding, the location of welds is a key issue. The fixture method adopted in the 1970s is prone to problems such as model loss, rough workmanship, and inaccurate positioning. In the early 1990s, technical solutions such as the use of ultrasonic receivers appeared, but they did not achieve further development due to serious environmental interference factors. In recent years, with the rise of machine vision technology and structured light technology, At present, most recognition welding equipment uses the first collection of weld images and then combines with the scale-invariant feature transform (SIFT) feature point monitoring method proposed by Lowe[1] or the Harries corner detection algorithm[2] to extract feature points, then based on the existing experience and working geometric relationship algorithm. However, this kind of algorithm has poor generalization ability and is easily interfered. It has strict requirements on the actual working environment of the site.

With the rise of deep learning and visual technology, models represented by the deep convolutional neural algorithm network such as Alex-Networks proposed by Krizhevsky et al.[3] and Vgg-net proposed by the Oxford University Machine Vision Group[4] adopts a weight sharing method similar to the biological nervous system. Zhou Zhihua[5] believes that deep learning neural network has simple input characteristics and strong nonlinear representation ability. Based on the above analysis, it can be seen that deep learning technology can be applied to the identification and positioning of industrial welds.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0048

Faster R-CNN is one of the best comprehensive performance methods in the current target detection algorithms based on the regional convolutional neural network series, but its detection accuracy for multi-target and small target situations is not high. Literature[6] proposed a regional convolutional neural network model, which uses a selective search method to select several candidate regions in the image to be detected, and uses a deep convolutional neural network for high-level feature extraction. Then use multiple SVMs to classify the features to complete the target detection task. Literature[7] proposed a fast RCNN (Faster-RCNN) model in order to improve the detection accuracy and detection speed of the RCNN model. The core idea is to first use the RPN network to filter out the region of interest, and then train on this basis. This method is characterized by high accuracy and slightly slower speed. This paper analyzes the basic working principle of the industrial weld position recognition system and the fast regional convolutional network model, and proposes an improved algorithm based on the Faster R-CNN model, introducing the number of regions adjustment layer and residual network, use this model to conduct data experiments on the HF data set and working platform. Comprehensive experimental analysis results show that this method can accurately identify the location of the weld.

2. Weld Recognition System

The structure of the industrial weld recognition system is shown in Figure 1. The weld image is uploaded to the host computer or computing unit through the vision sensor, and the image processing algorithm is used for analysis after preprocessing. The coordinates of the recognized start and end points of the fillet or lap weld are converted into three-dimensional coordinates and sent to the terminal. The terminal mobilizes the drive control motor and the welding gun to weld the weldment, and finally completes the workpiece welding task[8]. Fan Dejin et al.[9] used the least square method to fit industrial welds. Zhou Hongming et al.[10] used and analyzed the Hough transform method to identify the performance of industrial welds. Xu Hao[11] adopted SIFT feature point extraction method to identify industrial welds.

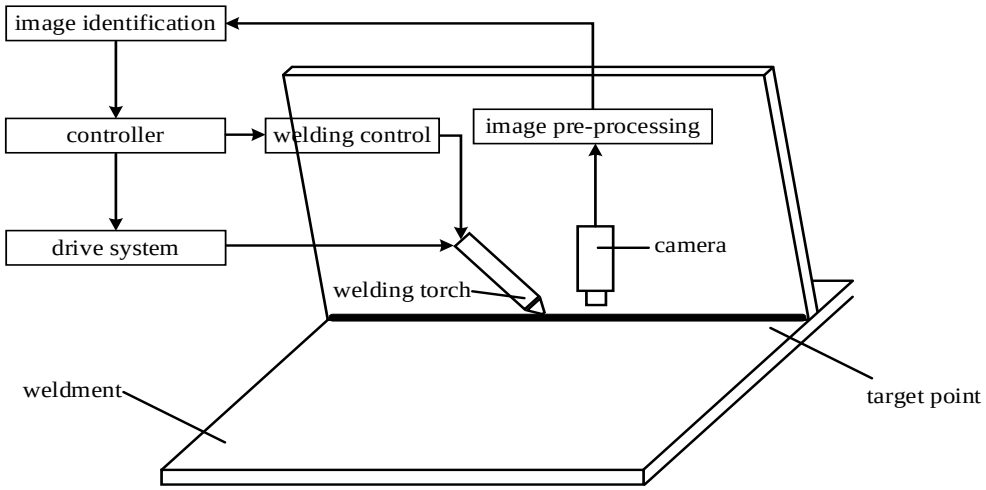

Fig. 1 Welding control recognition system

3. Faster R-CNN Algorithm and Improvement

Faster R-CNN is a classification-based target detection algorithm, and it is one of the popular target detection frameworks[12]. The algorithm framework is shown in Figure 2, using VGG16 as the feature basis to select the network, and adding convolutional layers, pooling layers and other structures on top of this to obtain a feature map[13]. Then a more accurate candidate area is generated through the RPN network, and finally ROI pooling is used to compare the output results of the feature map and the RPN network.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0048

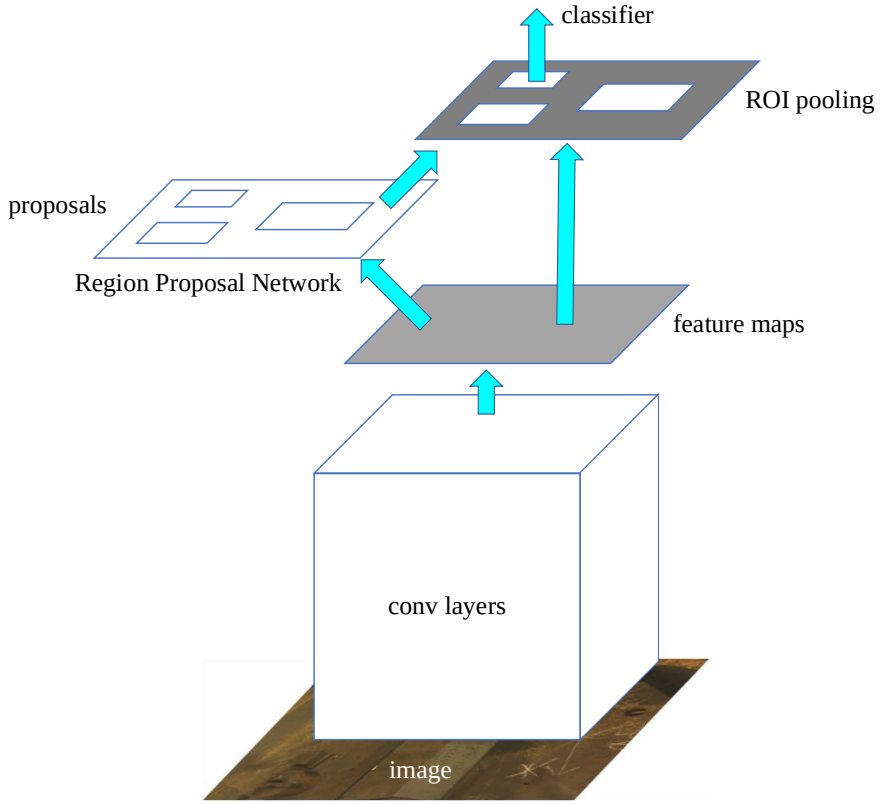

Fig. 2 The basic framework of the Faster R-CNN algorithm

The core idea of Faster R-CNN is to generate a number of candidate regions through pre-established anchor frame styles, and then perform subsequent operations such as judgment. The training image gets the feature map through the VGG16 network, and the original image is convolved several times to get the feature map. Each pixel on the feature map corresponds to a certain area on the original image.

In order to capture all the targets to be detected, the one-to-one mapping format does not meet the requirements. The following expansions are made to the receptive field area mapped on the original image: 1) Added three types of regional aspect ratios, namely $1:1$, $2:2$, $2:1$; 2) The size of the additional area is divided into three types: $512,256 and 128$. Each pixel on the feature map corresponds to 9 candidate regions in the original image. The number of candidate regions on the original image is 9 times the number of feature images. Although these candidate regions are sufficient to capture all the targets to be detected, they still cause the following problems: 1) A considerable part of the candidate regions are not of high quality; 2) There are too many candidate regions, which makes the computational overhead of training and detection too large. Therefore, using the technique of non-maximum suppression, all candidate regions are scored, and the candidate regions with higher overlap are discarded. Select the first 2000 candidate regions with higher scores for subsequent operations.

3.1 Faster R-CNN Improved Algorithm

This article mainly improves the performance through two aspects: 1) Improve the speed of the classic model. The classic method is to select 2000 candidate regions through maximum suppression. The method of generating candidate regions in this paper is to adopt adaptive feedback adjustment. Through feedback adjustment during training, the amount of candidate regions can be dynamically changed from 300 to 2000, effectively reducing training time. 2) Improve the recognition of the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0048

classic model, improve the feature extraction network, and introduce the residual network[14] to make the network deeper and the extracted features more abstract. The original 16-layer convolutional layer is reconstructed into 58 layers, which improves the accuracy of target detection. The above method can effectively reduce the training time and increase the average accuracy value. So that the improved model has improved recognition rate and training speed. The framework of the improved algorithm is shown in Figure 3. Subsequent improvements are based on the detection of specific targets (industrial welds). In order to meet the inspection of industrial welds in the actual workshop, the following improvements have been made, such as multi-scale training.

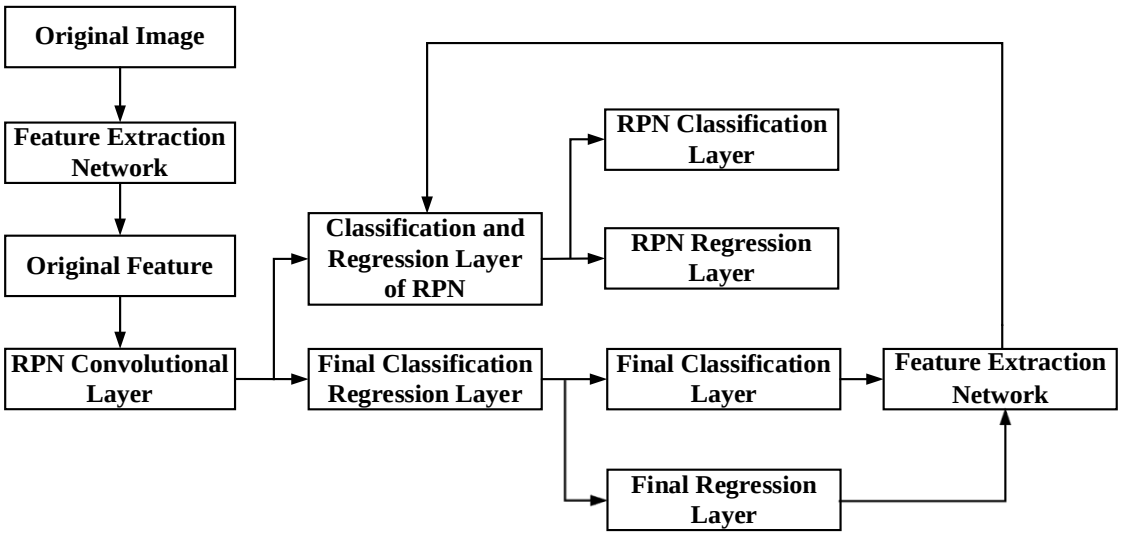

Fig. 3 Improved Faster R-CNN algorithm framework

3.2 Feature Extraction Network

The feature extraction network transforms the original image into a feature map rich in feature information. In order to obtain feature maps with more information, it is necessary to use a more excellent performance convolutional neural network to extract more abstract and complex information. The original feature extraction network of Faster R-CNN is VGG16. VGG16 is divided into 5 large sections as a whole, and the feature map is obtained through repeated convolution and pooling. The Faster R-CNN original feature extraction network has 16 layers. Through experiments, it is found that the accuracy can be improved by simply superimposing the convolutional layer. When the number of model layers is increased to 20, the experiment found that the average accuracy value not only did not improve, but decreased. In order to improve this problem, the model depth is further deepened, and the residual network is introduced.

Residual learning is not a simple stacking of convolutional layers, it is explicitly mapped in the residual layer[15]. It is difficult to train deep neural networks. Using the deep residual module when building the network can reduce the amount of data for deep network training and complete the training of deeper networks.

Assuming that the input of the residual learning module is x and the function to be fitted is mapped to $H(x)$, another residual mapping can be defined as $F(x)$, and $F(x) = H(x) - x$, then the original function mapping $H(x) = F(x) + x$. Experiments show that it is much easier to optimize the residual mapping $F(x)$ than to optimize the original function mapping $H(x)$ [16]. Direct mapping is just an identity mapping, without any other parameters, and does not interfere with the complexity and computational complexity of the overall network. Use this shortcut to build a richer convolutional neural network. The residual network is shown in Table 1. Convolutional layers two, three, four and five all adopt 3 convolution kernel modules, the convolution kernel 1 style is 1×1, the convolution kernel 2 style is 3×3, and the convolution kernel 3 style is 1×1.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0048

Table 1. Feature extraction network structure

Convolutional layer name	Convolution kernel dimension	Quantity
Convolutional layer one	64	1
Convolutional layer two	256	4
Convolutional layer three	512	5
Convolutional layer four	1024	7
Convolutional layer five	2048	3

The overall network structure is divided into 3 convolutional blocks, and each convolutional block is divided into 5 large convolutional layers. The first 7×7 convolutional layer performs dimensionality reduction, then passes through 4 convolutional layers, and finally passes through 1 convolutional layer to restore. A "shortcut" is used between the first convolutional layer and the second convolutional layer, and a "shortcut" is used between the third 3×3 convolutional layer and the last convolutional layer, Then use a "shortcut" between the first convolutional layer and the last convolutional layer. The structure has 58 layers. This network structure effectively improves the depth of the model.

3.3 RPN Network and Training

After the emergence of the Faster R-CNN network, most target detection networks used regional suggestion networks to speculate and identify the target location[17]. The core idea of Faster R-CNN is RPN, which uses convolutional neural networks to generate candidate regions. In order to select all the targets to be detected as much as possible, a large number of candidate regions need to be generated. For example, the original image of 1200*800 pixels is passed through the feature extraction network to obtain an 80*60 pixel feature map (the anchor frame style is set to 9 different types of candidate region modes), and the original feature image is traversed to generate 80*60*9 candidates area[18]. So many candidate regions cannot all be used for subsequent training. Therefore, the non-maximum suppression method is used to suppress high overlap candidate regions, and 2000 regions with higher quality are selected for training. The loss function of RPN is defined as:

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}} + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \tag{1}$$

$$L_{reg}(t_i, t_i^*) = R(t_i - t_i^*) \tag{2}$$

$$R(x) = \begin{cases} 0.5x2, & |x| < 1 \\ |x| - 0.5, & others \end{cases} \tag{3}$$

In the formula: i represents the serial number of the anchor point of each batch of data. p_i represents the probability of the $i-th$ target candidate region. When the corresponding labeled data is a positive sample, p_i^* is 1, otherwise, p_i^* is 0. t_i indicates the coordinates of the prediction candidate area. t_i^* represents the data coordinate labeled by the positive sample. $L_{cls}(p_i, p_i^*)$ means position regression loss. $p_i^* L_{reg}(t_i, t_i^*)$ means that only when there is a positive sample, it is counted as a regression loss. λ represents the weighting factor.

This paper introduces the NP (number of proposals) layer in the training, and automatically adjusts the number of candidate regions selected by RPN during training.

$$N_{p_i+1} = \begin{cases} N_{p_i}(1+\mu_1) & L_i \geq 2L_{i-1} \\ N_{p_i} & 0.5L_{i-1} < L_i < 2L_i \\ N_{p_i}(1-\mu_2) & L_i \leq 0.5L_{i-1} \end{cases} \tag{4}$$

In the formula: i represents the training number every N times. N_{p_i} represents the number of candidate regions used from the N_i th training to the $(N+1)_i$ th training. L_i represents the mean value of the regression loss from the N_i th training to the $(N+1)_i$ th training. μ_1 represents the penalty factor, μ_2 represents the reward factor.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0048

The NP layer is introduced in the training to adjust the training result synchronously and feedback. Calculate the mean value of the regression loss every N times of training. After a blank control experiment (fix N_{p_i+1} and take different values for experiment), it is obtained that every interval N times L_i is reduced by half and self-increased by 1 time, which is a reasonable change jitter interval. Exceeding requires feedback adjustment. Set the upper and lower intervals of the number value of the candidate area, and let the candidate value change adaptively in the interval of [300-2000].

In the actual application process, various target image sizes vary greatly. The original Faster R-CNN will fix the size of all training images, resulting in poor generalization ability for detection of different sizes[19]. This question uses multi-scale training. Before sending the image to the network, under the premise of ensuring the original ratio, the image is randomly adjusted in size, and the shorter side is one of 480, 650, or 850. Then choose one of the three random scales and send it to the network model for training. This operation is shown in Figure 4. Experiments show that multi-scale training can promote the network to learn the characteristics of various sizes of the target image, so that the network model has a certain degree of robustness to the size of the target image.

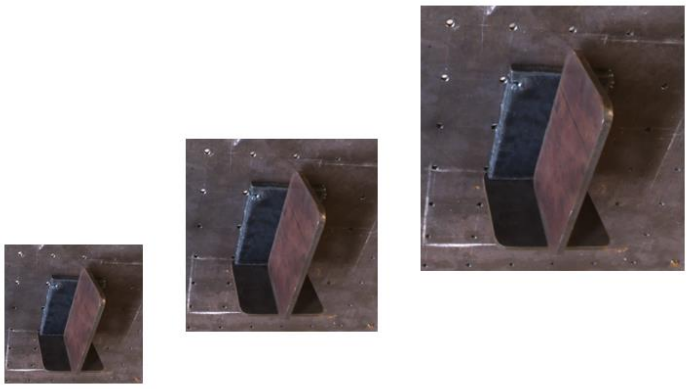

Fig. 4 Multi-scale training

4. Experiment Analysis

The experiment was carried out under the Ubuntu16.04 operating system on Intel(R)Core(TM)i7-9750 CPU, 8GB RAM, and RTX2060 graphics card. First, explain the labeling process and recognition labeling. Secondly, build a deep convolutional network model for training, then use the improved Faster R-CNN algorithm model, and finally test the model on the self-made HF data set and analyze the performance indicators.

4.1 Data collection, Annotation and Data set establishment

The MV-GE500C-T-CL industrial camera from Mindvision was used to collect 10,000 images of lap and fillet steel welds, as shown in Figure 5. In order to facilitate the training of the convolutional neural network, each image is cropped into a square image with the size of 224 pixel 224 pixel with the marked line as the center. Among them, 8,000 were used as the training set and 2,000 were used as the test set.

Fig. 5 Weld image acquisition equipment

ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0048

Mark the collected weld images as samples, and the marked content is the image coordinates of the beginning and end points of the weld. As shown in Figure 6, the collected original weld image is image_file. The row coordinate of the starting point is 565 (represented by r_{t1}), the column coordinate is 553 (represented by c_{t1}), the row coordinate of the end point is 554 (represented by r_{t2}), and the column coordinate is 830 (represented by c_{t2}), then store the row and column coordinates of the beginning and end points in the label file label_xml to complete the labeling of a single weld image.

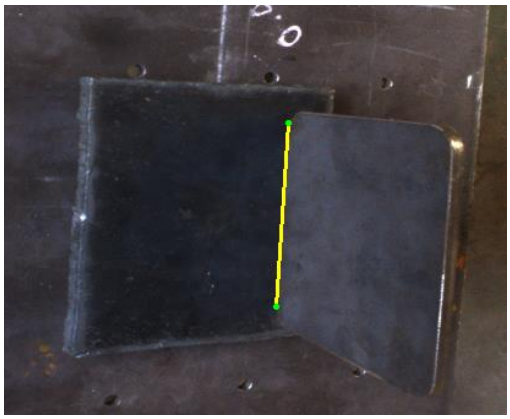

Fig. 6 Labeling of Workpiece 1

This article uses the training set and test set of the self-built HF data set, and the experiment uses the Pytorch framework to implement the convolutional neural network model, The parameters such as random deactivation, maximum iteration value, batch size in Faster R-CNN have a greater impact on the average accuracy value (mPA). To obtain a better output, these parameters need to be optimized. In the experiment, the maximum number of iterations is 70,000, the batch size of the RPN network is 256, and the random deactivation value is selected as 0.5.

Experiment 1 is the performance of the target detection network with the adjustment layer of the number of regions in terms of average detection accuracy and speed. Through many experiments, the speed is increased by about 16% on average, with almost no loss of accuracy, which saves a lot of money. The results of experiment one are shown in Table 2.

Table 2. Analysis of detection results in the same network with different numbers of candidate regions

Network Models	Number of Candidate Regions	Average Detection Accuracy (%)	Total Time (/s)
VGG16	2000	73.54	53688
VGG16	1000	73.55	51576
VGG16	500	73.19	47510
VGG16	250	72.63	46570
VGG16	100	71.58	45085
VGG16	10	57.66	44175
VGG16+NP	Trained	73.59	45721

Experiment 2 tested whether increasing the number of regions adjustment layer on the most advanced target detection model can increase the training speed, and whether the number of selected candidate regions is the best number. Most of the experiments in this article are based on Faster R-CNN. From Table 3, we can see that in the latest series of Faster R-CNN, the performance in Mask-RCNN still has a strong speed improvement. Experiments show that in the Faster R-CNN series, adding a candidate region adjustment layer has the effect of generally increasing the speed. From the second experiment, it is of great significance to add the region number adjustment layer in the Faster R-CNN series.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0048

Table 3. Analysis of the detection results of the adjustment layer of the number of added regions in different networks

Network Models	Data Set	Total Time(s)
Faster R-CNN	HF Dataset	53688
Faster R-CNN+NP	HF Dataset	45811
Mask R-CNN	HF Dataset	45277
Mask R-CNN+NP	HF Dataset	38714

The third test combines two improvements to the target detection effect of the weld position. As shown in Table 4, multi-scale training can improve the robustness of the network model, and adding a residual network can improve the reliability of the network model. Experiments show that these two improvements have achieved good results for the detection of weld position models.

Table 4. Improved neural network model's detection result analysis

Network Models	Detection Time (frame/s)	Precision (%)
VGG16	0.653	85.13
VGG16+ Multi-scale training	0.677	85.13
VGG16+ Area number adjustment	0.652	86.40
VGG16+ Two improvements	0.678	86.12

Through the analysis of the data results of Experiment 1 to Experiment 3, it can be concluded that adding an adjustment layer for the number of candidate regions on the target detection model can effectively improve the quality of the model.

The results of experiment four verify the effect of the improved feature extraction network and different feature networks in the target detection. As shown in Table 5, the reconstructed 58-layer residual network is 1.2% more accurate than the popular residual network.

Table 5. Analysis of the average detection accuracy of different networks

Network Models	Data Set	Precision (%)
ZF+RPN	HF Dataset	58.6
VGG16+RPN	HF Dataset	71.6
ResNet50+RPN	HF Dataset	74.0
Res58+RPN	HF Dataset	74.2

From the analysis of the experimental results, it can be seen that the residual network can effectively optimize the degradation problem of the neural network. When the VGG16 network generally deepens the convolutional layer with a 19-layer model, the map can be increased by 1%, As we continue to deepen the model, the accuracy of the map drops drastically. The residual network method can basically solve the problem of network degradation. After adopting the fast adaptive algorithm of the NP-layer feedback mechanism, the speed is increased by 16% and the effectiveness of the model is improved.

5. Conclusion

The seam recognition technology based on visual sensing focuses on image processing. Traditional algorithms are susceptible to various interferences such as light and noise, and it is difficult to expand the scope of application. The improved Faster R-CNN algorithm proposed by this question mainly focuses on two aspects of research:

1) Introduce the optimization model of the adjustment layer for the number of candidate regions, and optimize the model for the target detection of the weld position. Through the introduction of adaptive feedback of the number of candidate regions for dynamic adjustment, to achieve the purpose of increasing the speed and improving the effectiveness of the model; 2) Improve the reliability of the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202106_7(6).0048

model, use the advantages of the residual network to reconstruct and optimize the underlying feature extraction, and the accuracy is increased by 3.7%.

Aiming at the real-time and accuracy requirements of weld target detection, based on the convolutional neural network, a variety of framework models are built for training. By comparing the performance of each network, find the network suitable for welding seam identification. Compared with the traditional target recognition algorithm, the target recognition algorithm based on residual error can avoid the error caused by artificially extracting target features, and the recognition accuracy has been greatly improved. The experimental results show that the improved algorithm proposed in this paper has the advantages of high recognition rate, fast speed, small model and diversified sample recognition, and can be deployed in the recognition of weld positions in industrial welding sites.

References

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

[2] Zhang G G. Summary of corner detection technology[J]. Digital Technology and Applications, 2013, (4):157.

[3] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]// Advances in Neural Information Processing Systems,2012:1097-1105.

[4] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint: 1409.1556,2014.

[5] Zhou Z H. Machine learning[M].1st ed. Beijing: Tsinghua University Press,2016:213-215.

[6] Sang J, Guo P, Xiang z, et al. Vehicle detection-based on faster-RCNN[J]. Chongqing Daxue Xuebao/ Journal of Chongqing University, 2017, 40(7):32-36.

[7] Shi Yongsheng. Improvement of AODVjr routing algorithm in wireless sensor networks [J]. Electronic Technology and Engineering, 2013(22): 63-64.

[8] Sun W X. A weld seam recognition and tracking system of welding robot based on machine vision[D]. Qingdao: Qingdao University of Science&Technology,2018:13-17.

[9] Fan D J, Yang L X, Ding L, et al. Weld line extraction based on improved least square method[J]. Hot Working Technology, 2018, 47(15):217-220,224.

[10] Zhou H M, Lu J F, Lü J S, et al. Research on weld defect type recognition method based on multi-scale texture features[J]. Electromechanical Technology,2018(3):14-16.

[11] Xu H. Weld identification and trajectory planning based on machine vision[D]. Nanning: Guangxi University, 2017.

[12] Wang Xiaonan, Qian Huanyan, Tang Zhenmin. 6LoWPAN-based routing protocol for wireless sensor networks[J]. Application Research of Computers,2009, 26(10): 3881-3882+3887.

[13] Wang Wanguo, Tian Bing, Liu Yue. Research on power widget recognition of UAV inspection image based on RCNN [J]. Journal of Geoinformatics, 2017,,19(2):256263.

[14] Zhang Ke, Gao Ce, Guo Liru. Age estimation of face image in multilevel residual network under unrestricted conditions [J]. Journal of Computer-Aided Design and Computer Graphics,2018.

[15] Lu Yongshuai, Li Yuanxiang, Liu Bo. Haze monitoring of hyperspectral remotesensing data based on deep residual network[J]. Journal of Optics,2017,37(11):314-324.

[16] Li Wei, Zhang Xudong. Convolutional neural network-based super resolution reconstruction method for depth images [J]. Journal of Electronic Measurement and Instrument, 2017,31(12): 1918-1928.

[17] Jiang Shuai. Image recognition based on convolutional neural network[D]. Changchun: Jilin University, 2017.

[18] Zhang Ke, Gao Ce, Guo Liru, et al. Age estimation of face image in multilevel residual network under unrestricted conditions [J]. Journal of Computer-Aided Design and Computer Graphics, 2018,30(2) :346-353.

[19] Tang Xiaopei, Yang Xiaogang, Liu Yunfeng. Research on aircraft recognition based on deep convolutional neur-anetwork[J]. Electro-optic and Control, 2018,25(5):6872.