Research on Defect Detection Method for Steel Metal Surface based on Deep Learning

Xiaoyang Gai ,Peiran Ye,Jinglin Wang,Bingquan Wang
School of Information Engineering,Wuhan University of Technology
Wuhan, China
gxywhut@163.com, yorwuhan@163.com,1416833586@qq.com,wangbingquan@whut.edu.cn

Abstract—**In the process of modern industry, the surface defects of industrial products seriously affect the quality, safety, usability and aesthetics of products.Based on the method of deep learning, this paper takes steel surface defects in industrial parts as the breakthrough point, and mainly USES the convolutional neural network algorithm in deep learning to classify and detect steel surface defects.Firstly, industrial cameras were used to collect and pre-process the steel defect images to obtain relevant data sets. Secondly, VGG model was used to improve the network features to improve the recognition of defects and realize the classification and recognition of defects.Compared with traditional methods, this method has higher accuracy and efficiency.**

Keywords—surface defects;deep learning; convolutional neural network; classification recognition

I. Introduction

Industrial parts are widely used in a series of industries such as aerospace, machinery, electronics, and automobiles, and are indispensable components in various industrially manufactured products.In the highly competitive modern industrial production, the quality of industrial parts is directly related to the final quality of the product.In mechanical and automatic processing engineering, changes in the tool's trajectory during processing and the material's own characteristics, vibration, damage to the tool, and improper handling of the polishing process may cause appearance deformation, dents, and scratches on the surface of the machined part A series of flaws such as damage and unsatisfactory reflective characteristics. The surface flaws of industrial parts may not only make the appearance of the workpiece unsightly, but may also affect the performance of the workpiece.

With the continuous improvement of the processing requirements of modern industrial products, the detection of surface defects on machined parts has become an important method and means of product quality inspection and control. The variety and quantity of industrial parts urgently require enterprises to realize industrialized production automation.However, most companies currently use more traditional manual visual inspection methods. This method has a slow detection speed and cannot meet the high-speed manufacturing rhythm of industrial production. The sampling rate is low and the missed rate is high. , The workload is large, the labor cost required is huge, and there are large errors. The test results are easily affected by the subjective factors of the test staff. There is no consistent scientific guidance. Because there is no unified and reference evaluation mechanism, it also gives The formulation of standards brings inconvenience, so companies urgently need advanced technology and equipment for surface defect detection of industrial parts.

With the rapid development of artificial intelligence, it has become possible to use machine vision instead of manpower to solve such problems. Deep learning is a branch of machine learning and one of the key breakthroughs and researches made in the field of machine learning in recent years. Convolutional neural networks, as an important deep learning algorithm, have begun to be slowly applied to the field of image recognition. This paper is based on a convolutional neural network algorithm. Based on the VGG network model, by improving the VGG network structure, the detection accuracy of steel parts is improved.

II. The construction of the date

A. The data collection

In this study, the object of study is steel industrial parts, whose surface defects are not easy to distinguish with the naked eye. Therefore, it is necessary to use an industrial camera for image acquisition, and at the same time to overcome the interference of external factors such as light sources and set an appropriate focus And other parameters to obtain pictures with high imaging quality, and also to ensure that the acquisition environment is clean and the acquisition platform is stable, etc., to reduce the impact of external adverse factors on image acquisition.The CCD industrial camera used in this study covers 300,000 to 14 million pixels, and uses a more stable and versatile Gigabit Ethernet network for transmission. It has clear and stable image quality and supports external trigger input control acquisition and signal output. The industrial lens used with it has the characteristics of variable magnification, large depth of field, no optical distortion, high definition and high contrast. The focal length method is adopted, and the measurement error is small, which is suitable for applications such as highly accurate detection, microscopic magnification, and size measurement.Through the use of cameras and lenses, you can achieve image sample magnification and high-quality collection. In addition, you need to set up a collection platform, select a suitable light source and camera stand, and solve the problems of external light interference and shooting stability, making the pictures more Be clear and get high-quality data sets. The types of defects in the collected steel sample set are divided into rubbing, bottom leakage, bumping, and convex powder. Finally, the data samples are stored on the computer to complete the

978-1-7281-4323-1/20/$31.00 ©2020 IEEE

data set establishment. The process of collecting the data is shown in Fig.1.

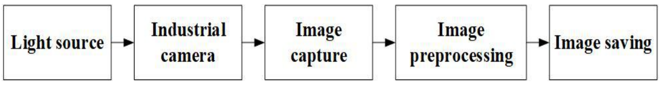

Fig. 1. The flow chart for collecting data.

B. *The enhancement of date*

Deep learning needs to learn based on a large amount of data. If the number of data is too small, overfitting will occur. It is often difficult to obtain a sufficient amount of data in the industrial production process, but the data set can be expanded by means of data enhancement. Therefore, the number of data sets is greatly increased, thereby reducing the occurrence of overfitting during the convolution training process, and the network generalization ability is further enhanced.

At present, data enhancement mostly enhances the read-in images, such as performing flipping, panning, and other operations, so that the data set can grow non-linearly. Commonly used methods are rotation, which changes the orientation of objects in an image by angularly rotating the original image. You can also use the flip transform to reverse the image. Zooming is to enlarge or reduce the original image proportionally. A translation transformation is a translation of the position of an image. Image filtering can change the blur degree of the image. Noise perturbation is to amplify the data set by adding random perturbations to the image. While performing data enhancement, you should also perform category balance, that is, to enhance the different types of images to different degrees, so that the number of various types of samples in the data set are similar. A type-balanced data set is more conducive to model training.

Due to the limitation of the number of samples and laboratory conditions, the number of sample sets obtained is not very large. There are 300 images in total, of which 34 are scratched images, 160 are missing images, 30 are concave images, and 76 are convex images. However, this number of data sets is far from sufficient for the training of deep convolutional networks. Therefore, we use the image enhancement method to enhance each type of image to expand it to about 300. The main operations include image rotation, translation, and miscutting. Finally, a defective sample set with a sample size of about 1400 is established.

III. DETECTION ALGORITHM BASED ON CONVOLUTIONAL NEURAL NETWORK

A. *Convolutional neural network*

Convolutional neural network is one of the most representative algorithms for deep learning. It is a hierarchical network structure containing convolution operations. Convolutional neural networks have great advantages in image processing because they share convolution kernels, are good at processing high-dimensional data, and feature automatic location. The so-called deep learning is actually to build a deep neural network model containing multiple hidden layers, and with the support of the computer's powerful computing power, it uses a large amount of training data to continuously optimize network parameters, making the constructed deep learning neural network better feature expression and classification capabilities.

B. *VGG network model*

VGGNet is a convolutional neural network model first proposed by Oxford University and Google Deep Mind. It uses a small convolution kernel size instead of a large convolution kernel size to increase the depth of the network, and repeatedly stacks multiple convolutional layers and maximum pooling layers. Reduce the input image size, maintain the translation invariance of the neural network, and use the pre-trained data of the specific layer to initialize the parameters. Compared with the AlexNet network structure, it not only improves the recognition function of the discrimination function, but also reduces unnecessary parameter.

The prominent feature of VGG is that it uses a small 3 × 3 convolutional layer, which increases the network depth and effectively improves the model's effect. Moreover, VGGNet has a good generalization ability for other data sets. The VGGNet model studies the relationship between the depth of CNN and its performance. By repeatedly stacking a small convolution kernel of 3 × 3 and a maximum pooling layer of 2 × 2, a CNN with a depth of 16-19 layers can be constructed.

The basic algorithm model used in this study is the VGG16 model. VGG16 has 16 layers, 13 convolutional layers and 3 fully connected layers. After the first two convolutions with 64 convolution kernels, one pooling is used, and the second time with 128 convolution kernels. After that, pooling is used, and the three 512 convolution kernels are repeated twice, and then pooled, and finally three full connections are performed. The VGG16 network model is shown in Fig. 2.

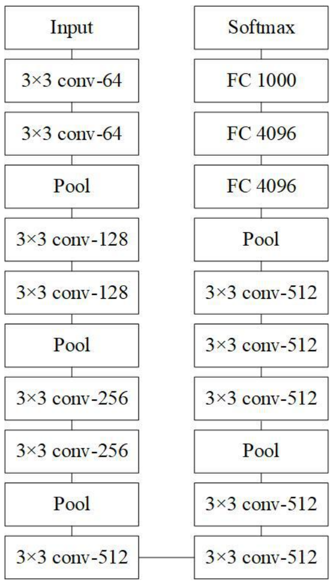

Fig. 2. VGG16 network model diagram.

C. *Improvement of network structure*

Although VGG16 is a model with outstanding classification performance in convolutional neural networks, it was found in experiments that the network structure has a large loss in image feature extraction, which is insufficient, affecting the final target detection results, leading to the recognition of parts targets The rate is not precise enough. With the continuous improvement of convolutional neural network models, deeper network models such as Inception and ResNet have appeared, and deeper network layers have improved performance. Therefore, combining the advantages of networks such as Inception and ResNet, the feature extraction network is improved, and the feature of the first three convolutional layers in the neural network can be used to extract features well, and the image features can be extracted more effectively. The improved VGG16 feature extraction network is shown in Fig.3.

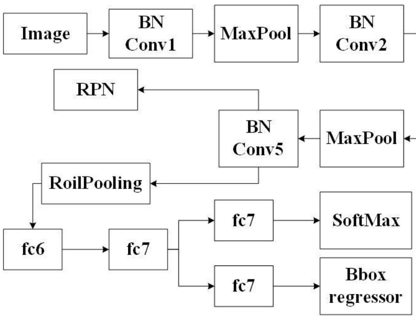

Fig. 3. The improved feature extraction network.

In the improved network structure, replace the Conv4 convolutional layer, borrow the idea of building a deep network from ResNet50, and the inner product method of Inception model split-transform-merge, and deepen the depth and width of the network at the same time to get more Strong expression skills.Increasing the depth and width of the network is the most direct way to improve the performance of the neural network. Since the Inception module is a network with a good local topology, parallel convolution operations are performed on the input images, and finally the different features obtained are stitched together. Therefore, when designing the network module, by combining the characteristics of ResNet and Inception, the Incepton model is referenced in the residual module to replace the convolutional layer in the residual connection to form a new structure, so that nodes learn between input and output. The difference mapping can avoid the fitting of input and output features, eliminate gradient dispersion and gradient explosion, and speed up the training of the network.

In the replacement module, in order to reduce the calculation amount, the 5×5 convolution is replaced by two 3×3 convolutions. First, dimensionality reduction is performed on the input by 1×1 convolution, then multiple 3×3 convolutions are used to transform, then channel merging is performed in series along the channel dimensions, multi-scale detection is performed, and finally a 1×1 convolution is used to achieve The channels are the same, and the addition of the residual module and the output linear vector is completed. The formula is shown in formula (1).When changing the network module, it can be designed according to a similar highly modular concept, so that the calculation complexity of the new module is similar to the conv4 module in VGG16, so that the network complexity can be maintained while increasing the network depth and width.

$$Z_{add} = \sum_{i=1}^{c} (X_i + Y_i) * K_i = \sum_{i=1}^{c} X_i * K_i + \sum_{i=1}^{c} Y_i * K_i \tag{1}$$

In the formula, X and Y are input channels, and the input and output channels in the algorithm are both 512. * is the convolution operation. In this module, since the 3×3 convolution layer is independent of the input and output, it is beneficial to the construction of the network. Therefore, the network module is designed by modifying the number of channels of the 3×3 convolution layer. The design module is shown in Fig.4.

The number of conv4 module parameters of VGG16 network and ResNet50 is 5898k, and each module of ResNet50 is about 983k parameters. Therefore, according to the principles of building similar modules, the number of channels of the 3×3 convolution layer is set to 172. In order to speed up the convergence speed, adding a normalization layer before each layer input is beneficial to correct the network output of the previous layer so that its average value is 0 and then input to the next layer. The improved feature extraction network can take advantage of different networks, improve feature utilization, and further enhance performance.

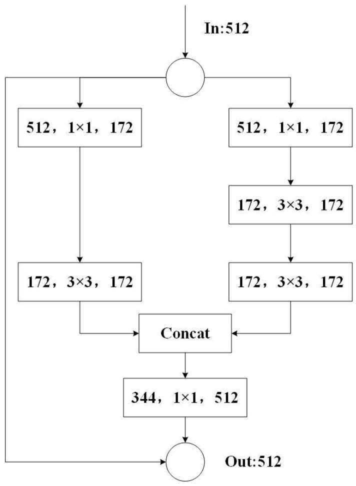

Fig. 4. The new design module.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Performance evaluation index

The performance evaluation index adopts average precision *AP*.It is often used as a measurement standard, mainly including two values: precision and recall. Precision is precision, which means precision, and represents the proportion of samples belonging to this category in samples that are predicted to be positive in different categories. The formula is:

$$precision = \frac{TP}{TP + FP} = \frac{TP}{N} \tag{2}$$

In the experiment, four types of samples, including scratch, bottom leakage, concave impact and convex powder, were used as the detection targets to classify and predict the types of defects. *TP*(True Positive) indicated the number of defects correctly recognized. *TN*(False Negativa) represents the number of unrecognized defect samples. *FP*(FalsePositive) error identifying flaws. The number of samples and recall recall rate, on behalf of the correct number and test set to detect the target all the ratio of the number, the formula is:

$$recall = \frac{TP}{TP + FN} \tag{3}$$

The formula of average accuracy is:

$$AP = \int_0^1 p(r)dr = \sum_{k=1}^{N} p(k)\Delta r(k) \tag{4}$$

From the above, *AP* is an integral of precision and recall.That is, the product of precision and recall for each threshold value is calculated respectively, and then the product value under all thresholds is accumulated.By synthesizing sample categories, the model can be evaluated by means of the average precision rate mean *MAP*, which represents the mean of the average accuracy of all sample categories in the model. The calculation formula is as follows:

$$MAP = \frac{1}{N}\sum_{n=1}^{N} AP \tag{5}$$

B. Test results

The experimental hardware platform operating system is Win10 64-bit, CPU Intel Core (TM) i7-4790 3.60GHz, 8GB memory. The entire experiment was developed based on the Tensorflow deep learning framework, the programming language was Python, and GPU-accelerated training was performed using a GTX1060 graphics card.

The experiment uses the collected steel parts dataset as a training sample. The model first optimizes the algorithm's weights. The momentum is set to 0.9, the weight attenuation factor is 0.05, the batch size used in each step is set to 128, and the initial learning rate is 0.01. When the number of batch processing reaches 30,000, the learning rate is adjusted to 0.0001; when the number of batch processing reaches 50,000, stop training to obtain the model, and save the trained model h5 file.

During the training of the basic model of defect detection, it was found that the loss of the network did not change after about 50 cycles of training, that is, the defect detection model had converged. Therefore, in the subsequent improvement of the network structure for training, the training period is set to 50.The training model of the VGG improved network structure is transferred to the test data set, where the test set includes 400 defective samples (100 samples of four types of defective samples). After testing the test set, the network evaluation results obtained are shown in TABLE I.

TABLE I. *TEST RESULTS*

Algorithm model	*AP*	*recall*	*MAP*
VGG foundation model	63%	67%	65%
The improved network model	74%	80%	77%

As can be seen from the table, the average accuracy rate of the improved network structure model based on VGG is significantly improved compared to the basic network model, which shows that the algorithm has a fairly obvious effect after improvement, and further proves that the method is effective in defect detection. It was also found in the experiment that due to the large proportion of bottom leakage, the detection results are relatively close to this situation. Therefore, in future research, it is necessary to increase the number of defective data sets for training the network, while ensuring the data concentration The number of various types of defects is relatively average in order to increase the average accuracy and network detection capabilities.

The total loss of the defect detection network (total_loss) is composed of classification loss, regression loss in the regional recommendation network, and classification loss and positioning loss in the main network. It is necessary to visualize the loss curve on the training data set in time during the training process, which can guide the adjustment of various parameters in the defect detection model, including the number of iterations, learning rate, weight attenuation coefficient, and structure optimization.

The number of training cycles of the VGG improved network structure used in this project is 50. Each cycle will be trained on all training data sets, and the various types of loss values of the network will be recorded every 50 steps. Fig. 5 shows the total loss curve of the training process. As the training progresses, the model training accuracy continues, the loss function gradually decreases, and the loss value decreases greatly at the beginning of training, indicating that the learning rate is appropriate and there is a gradient descent After the training reaches a certain stage, the loss curve becomes stable, indicating that the trained model starts to converge.

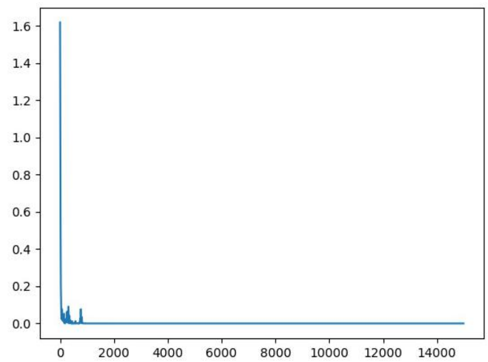

Fig. 5. The large-scale network survivability chain.

V. CONCLUSIONS

With the development of industrial automation, rapid and accurate quality inspection of industrial products is of great significance.With the development of deep learning, industrial detection algorithms have been greatly improved in both speed and accuracy compared with traditional algorithms.As an important product of industrial production, steel products will have a variety of defects in the process of production and processing.This paper is based on the deep learning detection algorithm to identify defects, improve the detection efficiency, in industrial production has a certain degree of universality.

This paper mainly studied the VGG improved network based model of steel surface defect classification and detection, through introducing convolution in deep learning neural network algorithm solved the problem of the traditional image processing in the narrow as well as the traditional artificial classification of strong subjectivity, low accuracy, the problem of low efficiency of classification, and for introducing deep learning other problems in the field of industrial provides reference.In the experiment, we first used the industrial camera to collect the steel images and set up the data set, and then enhanced the data to improve the recognition ability of the network and reduce the problem of overfitting.On the basis of Tensorflow deep learning framework, the VGG basic model is first built and then the network structure is improved, the number of network layers is deepened, and the ability to identify targets is improved.Through experiments, compared with the original network, the improved feature extraction network is more effective in feature extraction, and it is also found that the improved feature extraction network has higher accuracy in target detection.

In addition, in the defect detection process, due to the uneven distribution of sample types and quantities, it is necessary to continue to improve the sample and algorithm model in the later work to make the training model more reasonable and accurate, so as to further improve the accuracy of the training results.In the experiment, the deepening of network structure will lead to the problem of time wasting, long training time and slow response.In order to better apply to the industrial scene, in the future research, the network can be changed to a lightweight network for training, to better meet the needs of industrial detection.The improved detection algorithm means the improvement of detection technology, which has great practical significance to improve the level of industrial production.

ACKNOWLEDGMENT

This work was supported by the National innovation and entrepreneurship training program for college students.The project number is S201910497052.

REFERENCES

[1] D. Qi, P. Zhang, L. Yu, "Study on wood defect detection based on artificial neural network.." 2008 IEEE Conference on Cybernetics and Intelligent Systems(ICCI), Chengdu, China, pp. 21–24 Sept. 2008, IEEE, pp. 951-956.

[2] D. Dan, Y.Q. Zhang, "Research on Definition of Network Survivability," Journal of Computer Research and Development, vol. 43(Suppl.), pp. 525-529, 2006.

[3] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard，"Tensorflow: a system for large-scale machine learning" 12th Symposium on Operating Systems Design and Implementation (2016), pp. 265-283.

[4] J. Dai, Y. Li, K. He, J. Sun，"R-fcn: object detection via region-based fully convolutional networks " Advances in Neural Information Processing Systems (2016), pp. 379-387.

[5] H. Xu, L. Wang, S. Ni, "Application of Artificial Neural Network to Nondestructive Testing of Internal Wood Defects Based on the Intrinsic Frequencies," 2010 International Conference on System Science, Engineering Design and Manufacturing Informatization, Yichang, China, 12-14 Nov. 2010, IEEE, pp. 207–210.

[6] L.H. Wang, W. Qi, et al.Pattern recognition and size determination of internal wood defects based on wavelet neural networksComput. Electron. Agric., 69 (2009), pp. 142-148.

[7] N. Chen, X. Men, X. Han, et al. "Edge detection based on machine vision applying to laminated wood edge cutting process." 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), Wuhan, China, 31 May-2 June 2018, IEEE, pp. 449–454.

[8] Huibin Sun,Jiduo Zhang,Rong Mo,Xianzhi Zhang. In-process tool condition forecasting based on a deep learning method[J]. Robotics and Computer-Integrated Manufacturing,2020,64.