Defect Detection for Mechanical Design Products with Faster R-CNN Network

Ping Liu, Sheng Su, Xiaobing Gao, Hongfei Zheng, and Zilin Ma

School of Art and Media, Xi’an Technological University, Xi’an 710032, Shaanxi, China

Correspondence should be addressed to Ping Liu; liuping@mails.imnu.edu.cn

Received 12 February 2022; Revised 28 February 2022; Accepted 3 March 2022; Published 18 April 2022

Copyright © 2022 Ping Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The emergence of machine vision has promoted the automation of defect detection (DD) in the industrial field. Therefore, scholars at home and abroad have carried out a lot of research and exploration on the traditional visual DD method of mechanical design products. At the same time, this method has been widely used in the field of modern manufacturing due to its noncontact and fast detection speed. The traditional visual detection method is to use cameras, computers, and other equipment instead of people to detect the detected objects, although this method improves the production efficiency to a certain extent. However, this detection method is greatly affected by light, has a certain false detection rate, and has poor adaptability. The intelligent detection method based on deep learning developed on the basis of traditional vision is a further optimization of traditional visual detection methods. The rapid development of deep learning makes the advantages of visual DD more obvious.

1. Introduction

The research content of this paper is the surface defect detection (DD) of mechanical design products. Owing to the complex production process and uncertain environmental factors, it is difficult to avoid producing some inferior mechanical design products with surface defects. Defects on the surface of the workpiece directly affect the appearance quality of the workpiece, and the appearance quality of the workpiece determines the value of the workpiece and is directly related to the safety and stability during use. For example, as an important part of the DC motor, the tile-shaped permanent magnet provides a constant magnetic potential source for the motor. During the production process, defects such as cracks and pores may appear on the surface of the workpiece, which affects the service life and quality of the motor [1]. Metal bearings are very common and common parts in the machinery industry. They mainly play the role of supporting transmission parts, transmitting torque, and bearing loads. Different types of defects, such as roll printing, have a great impact on its performance and life and are likely to cause equipment failure [2]. The surfaces of rails and mechanical design products are prone to defects such as cracks, pits, and scratches, which will affect their bearing capacity, strength, and stiffness, and may even lead to safety accidents [3–5]. The surface of the light emitting diode (LED) chip is prone to defects such as scratches and cracks, which are directly related to the control of the quality of the LED chip and even cause the chip to fail to work properly [6]. Therefore, the surface DD of mechanical design products has become an indispensable and important link in modern industrial production. The traditional surface DD method of mechanical design products is often manual detection, but the traditional manual detection has the following shortcomings: (1) Subject to the influence of subjective factors, such as the experience, cognition, and thinking of technicians and objective factors such as illumination and environment, it is easy to cause the test results are subjective and nonstandardized. (2) Under the condition of repeated work for a long time, technicians will inevitably be mentally exhausted and visually fatigued, which may easily lead to missed detection and false detection in the test results. (3) It is difficult to detect small defects and defective workpiece surfaces with unclear color and texture. (4) When manual detection of surface defects of workpieces is based on contact, it may cause secondary damage to the workpiece to
be detected. (5) Due to the limited energy of human beings, the traditional manual detection work method is inefficient. There are many deficiencies in traditional manual inspection, and automatic nondestructive workpiece defect inspection is an inevitable trend of industrial development. Common physics-based nondestructive testing techniques include ultrasonic testing, infrared testing, magnetic flux leakage testing, radiographic testing, and visual testing [7–10]. Ultrasonic technology is suitable for various materials, such as metal and nonmetal, and it is difficult to detect workpieces with complex shapes. Eddy current testing has high precision for near-surface DD and is suitable for conductive metal materials. Radiographic inspection technology can detect internal defects in workpieces, but it is difficult to detect flat defects such as color spots. Automated nondestructive testing based on machine vision can be divided into three categories: traditional image processing, machine learning-based and deep learning-based algorithms. With the rapid development of the Internet, Big Data, and the improvement of hardware computing power, the development of artificial intelligence and deep learning technology has also been vigorously promoted, with many achievements in application fields such as image recognition, image segmentation, target detection, target tracking, and super-resolution [11, 12]. The background of surface defects of mechanically designed products is complex, and there are many types of defects. Various types of defects are represented in different forms, and the same type of defects is also different in form. It is difficult to capture all effective features through manual features. The deep learning convolutional neural network has a strong ability to extract image features autonomously, so it is meaningful to introduce deep learning into workpiece surface DD.

Therefore, the following work is done in this paper:

(1) The research status of domestic and foreign mechanical design product DD methods is introduced.

(2) Research and propose to use the superior faster region-based convolutional neural network (R-CNN) model for experimental verification of the sample set in this paper. This network is used for models in machine learning and object detection. By analyzing the structure of the faster R-CNN model, it is proposed that the model adopts the advantages of residual structure and feature fusion in faster R-CNN, and it is optimized from two parts of the feature extraction network (FEN) and RoI pooling structure.

(3) The algorithm of faster R-CNN is optimized, the methods of residual structure, feature fusion, and hole convolution are proposed for optimization, and the detection accuracy of several models is tested.

2. Related Work

Deep learning models have powerful representation and modeling capabilities. Through supervised or unsupervised training methods, they can learn the feature representation of objects layer by layer automatically and achieve hierarchical abstraction and description of objects. The large-scale research on applying deep learning to image classification and object detection began at ILSVRC in 2012, and Smirnov et al. [13] proposed a new CNN model, AlexNet, which achieved very low error rate on image classification tasks, which is nearly half of the error rate of the second-place-based method based on traditional methods. Owing to the success of AlexNet, many researchers began to focus on and improve the CNN structure. Zeiler and Fergus [14] reduced the size and step size of the first layer filter of AlexNet to form a ZFNet model. Yang et al. [15] then proposed the VGG network to explore the change in the performance of CNN with the increase of the number of layers when the total number of network parameters is basically unchanged. Szegedy et al. [16] proposed a new deep CNN model, GoogleLeNet. It only uses 12 times fewer parameters than AlexNet but has a lower classification error rate. GoogleLeNet adopts the inception structure, which combines multiple convolutional layers and pooling layers as the output of the inception module and uses a 1 × 1 convolution kernel to reduce dimensionality, which not only increases the depth of the network but also reduces the size of the network parameter. However, this structural gradient is unstable and difficult to train, so its improved model Inception-v2 appears. It adds batch normalization to normalize the output of each layer to keep the distribution of the input of each layer stable, so that the gradient is less affected by the initial value of the parameters [17, 18]. In 2015, He et al. [19] proposed a residual network ResNet with a depth of hundreds of layers. The number of layers is more than five times that of any previous successful neural network, and the image classification error rate on the ImageNet test set is as low as 3.57%. ResNet has strong versatility and achieved the best competition results at that time not only in image classification tasks but also in object detection and object localization tasks in the ImageNet data set and object detection and segmentation tasks in the MS COCO data set. At present, deep learning algorithms have become the mainstream method in the field of computer vision, showing excellent performance in various tasks and even surpassing human performance in some tasks. However, it can also be seen that its research is mainly concentrated in the theoretical field, and it still faces huge difficulties in practical application. Therefore, the application research in specific scenarios still has a broad space for development. The early surface DD based on machine vision mostly adopts traditional image processing methods. Zheng et al. [20] used genetic algorithm to learn morphological processing parameters for DD on uneven metal surface defects. The automatic pipeline DD designed by Wang and Su [21] performs image segmentation and feature extraction, uses K-means clustering, and combines the C45 decision tree to make an analysis decision. The advantage of this algorithm is that no additional space is needed for storing template images, and detection algorithms can be designed specifically. Zhao et al. [22] proposed a statistical average difference shadow method and an extensive grayscale correlation method for the detection of nontexture defects. Guo et al. [23] proposed a method combining Kirsch operator and
Canny operator to realize surface DD. This method can better suppress noise interference and more accurately locate the edges of surface defects. In recent years, deep learning has achieved rapid development in the field of computer vision, including natural image classification, face recognition, and object tracking. DD methods based on CNN have also been proposed by some scholars. Masci et al. [24] proposed a multiscale pooling method to detect defects on steel surfaces, which can accept images of different sizes as input. Natarajan [25] used a multilayer CNN to extract image features, realizes DD by transfer learning and also combines SVM and voting strategies to avoid overfitting. Wang et al. [26] proposed a method combining CNN with sliding window to locate defects in images. Chen et al. [27] proposed a DD system including three CNNs, two of which are used for defect localization and the other for defect classification. Ren et al. [28] used a pretrained deep learning model to classify defect images and adopted a heatmap-based segmentation method to obtain pixel-level defect information. There are relatively few studies on button surface DD. Most of these studies are based on traditional digital image processing methods, which are more sensitive to environmental changes and less reliable. Therefore, it is of great significance to study the application of deep learning methods in the DD of mechanical design products.

3. Method

In this section, completely describe the fast R-CNN algorithm, its working and accuracy of work. Then, it elaborates the feature extraction algorithm based on residual connection. Furthermore, it explains the feature fusion algorithm based on residual connection. In addition, talks about the optimization of feature fusion algorithm based on residual connection. Then, it explains optimization of pooling algorithm.

3.1. Fast R-CNN Algorithm. Compared with the traditional visual detection method, although the R-CNN method has improved the detection effect, there are still some problems. The number of effective regions obtained by the selective search method usually reaches more than 1000, which indicates that the neural network needs to repeat the calculation more than 1000 times. Therefore, this part of the calculation process is very time-consuming. In addition, training R-CNN requires training in multiple steps which is not only cumbersome but also slow to train. In the training process, all the features need to be saved which will occupy a large memory space. Since the final classification step is performed using a fully connected layer, the size of the input feature map is also required to be fixed. Ross Girshick proposed a faster and stronger fast R-CNN algorithm in 2015. The main detection process is shown in Figure 1. Compared with the original R-CNN, it is mainly improved in three aspects:

1. Convolution sharing: after using convolution to obtain the feature map from the entire input image, candidate regions are generated on the feature map, instead of generating candidate regions one by one like R-CNN and then extracting features for each region. The selective search method is still used in this process, but due to the advantage of convolution sharing, that is, only one convolution process is performed to obtain all the feature information of the entire image, so the amount of computation can be greatly reduced.

2. RoI pooling: the pooling idea in the SPPNet network is adopted, and feature scale transformation is performed by feature pooling. After adopting this method, images of any size can be input to meet the input requirements of the fully connected layer, making the training process more flexible and accurate.

3. Multitask loss: in this part, the classification is trained together with the regression network, instead of using the slower SVM classifier, the Softmax function is used for classification.

The fast R-CNN algorithm is based on the VGG16 network, and the training steps can be realized end-to-end, and the training speed, test speed, and detection accuracy have been improved. Although fast R-CNN has a better detection effect, in this algorithm, it takes 2 to 3 seconds to use selective search to generate regions, while it only takes 0.2 seconds to extract features using convolutional networks, so this kind of region generation method limits the performance of this algorithm and further optimization can be done. The faster R-CNN detection algorithm published in NIPS 2015 is an improvement on the fast R-CNN algorithm. As shown in Figure 2, its main detection process is similar to that of fast R-CNN, except that it will be used to generate the selection of candidate regions. The property search method is replaced by a region proposal network (RPN) structure.

The biggest advantage of faster R-CNN is that it proposes an RPN network, which maps the generated region to the feature map generated by the CNN through Anchor and realizes the connection between the two, and the detection speed and detection accuracy are further improved. Anchor is equivalent to many rectangular boxes with a certain size and aspect ratio on the image. Since the defects to be detected in the image are also marked with multiple rectangular boxes of different sizes and aspect ratios, therefore, Anchor is used as a strong prior in faster R-CNN, and then the Anchor is matched with the real defect, and then the defect classification and location are fine-tuned. As shown in Figure 3, the middle gray part is the calculation flow chart of the RPN algorithm. The detection is divided into two steps through RPN, and the proposal and preliminary positioning and classification are provided. The more accurate the proposal, the smaller the error of the subsequent redetection. The RoI is obtained by screening a large number of proposals generated by the predicted Anchor during training, and the proposal is directly used as the RoI during testing. RPN mainly includes five submodules:

1. Anchor generation: RPN corresponds to nine anchors for each point on the generated feature map. The nine Anchors are composed of three different
Convolutional Network
Classification
Regression
Input image
Generating regions on feature maps and pooling
Fully connected Classification and regression

Figure 1: Fast R-CNN detection process diagram.

Convolutional network
Input image
Feature map
RPN
Generated area
RoI pooling
Classification and positioning

Figure 2: Faster R-CNN detection process diagram.

Input image
Feature extraction network
RPN convolutional network
Generate proposal
Input tag
Calculate RPN loss
Generate Anchors
Filter proposals to get RoI
RoI Pooling
Computing the R-CNN ground-truth
R-CNN prediction network
R-CNN loss

Figure 3: Schematic diagram of faster R-CNN algorithm process.
area sizes and three different aspect ratios, corresponding to the defects that may appear in the original image.

(2) RPN convolutional network: by using a CNN, each generated Anchor is predicted to obtain its prediction score and prediction offset value.

(3) Calculating the RPN loss: this part only appears in the training process. Match the Anchor with the label to distinguish positive and negative samples, get the true value of the classification and offset, and calculate the loss of the predicted score and predicted offset value obtained in the previous step.

(4) Proposal generation: screen the anchors obtained through the RPN convolutional network to obtain a set of better proposals to prepare for the subsequent network.

(5) Screen Proposal to get RoI: screen the proposal obtained in the previous step to get the final RoI.

As shown in formula (1), RPN uses a multitask loss function, combines classification and localization losses, performs training in a unified manner, outputs the corresponding classification and frame position, and performs preliminary detection to provide the following R-CNN detection network. A high-quality proposal can improve the detection accuracy of the model. This module shares full-image convolutional features with the R-CNN detection network and saving time.

\[
L \left( \{ p_a \}, \{ f_a \} \right) = \frac{1}{N_{cla}} \sum_a L_{log}(p_a, p^*_a) + \frac{1}{N_{pos}} \sum_a p^*_a L_{pos}(f_a, f^*_a), \tag{1}
\]

where \( a \) is the index of the anchor; \( p_a \) is the probability that the anchor is predicted to be a defective target; \( p^*_a \) is the category label, which is 1 when there is a target, otherwise 0; \( f_a \) is the predicted value of the position; \( f^*_a \) is the label of the position \( a \) is the weight balance value; \( N_{cla} \) is the number of Anchors for classification; \( N_{pos} \) is the number of Anchors for position regression; \( L_{log} \) is the classification logarithmic loss; and \( L_{pos} \) is the position regression loss.

The classification loss \( L_{log} \) is the logarithmic loss function of the binary classification as

\[
L_{log}(p_a, p^*_a) = -\log(p_a p^*_a + (1 - p_a)(1 - p^*_a)). \tag{2}
\]

The location regression loss \( L_{pos} \) is defined as

\[
L_{pos}(p_a, p^*_a) = S(p - p^*_a), \tag{3}
\]

where \( S \) is the smooth \( L_1 \) loss function, and its calculation formula is as follows:

\[
\text{smooth}_{L_1}(x) = \begin{cases} 
0.5x^2, & |x| < 1, \\
|x| - 0.5, & |x| \geq 1.
\end{cases} \tag{4}
\]

It can be seen from equation (4) that the smooth \( L_1 \) function combines the first-order and second-order loss functions, which is more conducive to the convergence of the model during the training process. This is because when the difference between the predicted offset and the real value is large, the derivative of the second-order function is too large, which will cause the model to diverge and make it difficult to converge. Therefore, a first-order loss function with a smaller derivative is used when the input \( x \) is greater than 1. Compared with R-CNN and fast R-CNN, the faster R-CNN algorithm has obvious advantages in detection speed and accuracy, so this paper selects the faster R-CNN algorithm as the detection algorithm for the defect sample set in this paper based on the R-CNN algorithm series. The detection network model is established, and the experimental analysis is carried out.

3.2. Feature Extraction Algorithm Based on Residual Connection. If you want to obtain better defect characteristics of mechanical design products, it is necessary to improve the ability to extract defect characteristics of mechanical design products. The number of network layers can be superimposed by the network layer to obtain a deep network, which can improve the ability to extract the defect features of mechanical design products to a certain extent. The ResNet team has proved through experiments that the performance of deep network models will degrade: that is, at the beginning, the accuracy of the model will indeed increase with the deepening of the number of network layers, but when it reaches a certain depth, it will reach saturation, from this point on, the model will increase with the number of layers, and the effect of the model will become worse. Therefore, in view of the problem of model performance degradation caused by a large increase in the number of model layers, the ResNet structure is proposed. Through a network composed of some residual modules, the gradient disappearance of the model during training can be prevented and the degradation of model performance can be alleviated.

The so-called residual network structure refers to letting the CNN learn the residual mapping, rather than expecting each weight network layer to fully fit the underlying mapping. The cross-layer connection is implemented in ResNet, so that the passed gradient can bypass some layers and reach the input layer, which can alleviate the problems of gradient disappearance and model degradation, and make the model have better performance. The ResNet residual network composed of the residual structure can improve the problems that occur in the training and learning of the deep network, so that the network model with deeper network layers can be built and trained. The original faster R-CNN uses only 13 layers of convolution in VGG16 to extract the features of defects. The number of network layers is small, and the extracted semantic information is limited. Therefore, this section adopts the residual idea to optimize the feature extraction algorithm, that is, the ResNet50 network is used to replace the original FEN VGG16. In the ResNet50 structure, by using two \( 1 \times 1 \) convolutions before and after, and placing a \( 3 \times 3 \) convolution in the middle, the residual module composed of three different convolutions is called bottleneck. In the bottleneck module, first, the number of
channels of the feature map can be reduced using $1 \times 1$ convolution, and then after $3 \times 3$ convolution operation, $1 \times 1$ convolution is used to increase the number of channels, which can reduce the number of channels. The number of parameters also reduces the amount of computation.

3.3. Feature Fusion Algorithm Based on Residual Connection. Since the surface defects of mechanically designed products are characterized by multiscale and small objects, the adopted model needs to be suitable for the detection of such defects. Generally speaking, as the number of layers increases, the number of channels will also increase, and the scale of the feature map will become smaller and smaller. In the ResNet network, the down sampling rate doubles for each additional module and finally reaches a high down sampling rate of 32 times. The increase of the down sampling rate will lead to the phenomenon that small defects have less remaining feature information on the feature map, or even completely lose the feature information of small defects. To enhance semantics to improve detection accuracy, traditional DD models usually only perform subsequent operations on the last feature map of deep convolutional networks. The down sampling rate corresponding to the last layer is usually relatively large, such as 16 times and 32 times, and the resolution of the feature map is low, resulting in less effective information of defects on the feature map, in particular, the detection performance of small defects degrades sharply, a problem also known as multiscale problem. In order to realize the effective detection of multiscale defects, it is necessary to realize the extraction of multiscale defect features. For the multiscale problem, this section proposes several solutions:

(1) Feature pyramid structure based on image pyramid mainly scales the original defect image to scales of different specific sizes and then performs feature extraction and detection on each defect original image of different scales. In this method, the input image is scaled into multiple scales, and the defect images of different scales are extracted by convolution operation to obtain feature maps of different scales. This method is relatively simple, but it has many shortcomings. Owing to multiple image scaling and multiple feature extraction, the process is very time-consuming and computationally intensive. Therefore, although this method produces a multi-scale feature representation, which is beneficial to the detection of multiscale defects, it takes a lot of time and is not suitable for practical applications.

(2) Single feature map detection is the most commonly used detection method in detection algorithms, and the original model in this paper also uses this method to detect defects. This method extracts features layer by layer but only performs a single-feature mapping, and finally outputs the deep feature information. Although the deep feature map has strong semantics and is conducive to classification, due to the excessive number of down samplings, small-sized defects will be ignored, and small-defect features will be lost, resulting in a sharp drop in the performance of detecting small defects. In addition, since the deep feature map lacks the detailed information of the shallow layer, which is not conducive to the localization of defects, it is necessary to seek a better detection method.

(3) The pyramid feature hierarchy is to make multiple predictions on the extracted multilayer feature maps. Although large-size defects can be detected on deep feature maps with larger receptive fields, and small-size defects can be detected on shallow feature maps with smaller receptive fields, when small defects are detected on shallow feature maps, due to the semantic information insufficient, it will lead to poor classification of small defects, and in the detection of deep feature maps, due to the lack of shallow semantic information, it is not conducive to localization.

(4) Feature pyramid network structure uses the CNN to convolve again and again to form a pyramid shape with a hierarchical structure. The generated feature maps also have different scales, and these feature maps have high semantic information, so that multiple effective detection on a feature map. By considering the correlation between feature maps, a structure is designed through bottom-up convolution, top-down up-sampling, and horizontal connection addition and fusion. The features are combined with high-resolution, semantically weak features, and further fused using convolution operations to obtain multiple feature maps with different resolutions and semantic information. In DD, classification of defects requires a deep feature map with high semantics, but its resolution is usually very low, and when locating defects, detailed features are required, so the resolution of the feature map should not be too small. Therefore, through the feature fusion between the upper and lower layers, the high-level semantic information of the deep layer is transmitted to the bottom layer, and the semantic information of the shallow layer is supplemented, so that high-resolution and high-semantic features can be obtained. It can be seen from the aforementioned analysis that the feature extraction and fusion using the feature pyramid network has better advantages than other methods.

For the DD algorithm in this paper, RoI extraction needs to be performed on the feature map. For feature maps with different scales in four layers, RoIs with different sizes use different feature maps. The calculation method used by RoI to obtain the corresponding features for which layer of feature maps is specific is as

$$F = \left[ F_0 + \log_2 \left( \sqrt{\frac{lm}{224}} \right) \right]$$

where 224 is the image size used in ImageNet pretraining; $l$ and $m$ represent the length and width of the input RoI.
3.A. Optimization of Feature Fusion Algorithm Based on Residual Connection. Hole convolution means that there are some holes in the convolution kernel, and some elements are skipped for convolution. There is an additional hyper-parameter, which is the number of holes. The same $3 \times 3$ convolution can have the effect of $5 \times 5$ convolution or $7 \times 7$ convolution. It can be concluded that under the condition of the same number of parameters, the atrous convolution has a larger receptive field than the ordinary convolution and can cover a larger range. Assuming that the size of the convolution kernel of the atrous convolution is $c$ and the number of holes is $h$, the equivalent relationship between the atrous convolution and the ordinary convolution $c_1$ is shown in

$$c_1 = c + (c - 1) \times (h - 1). \quad (6)$$

When calculating the receptive field, you only need to replace the original convolution kernel size $c$ with $c_1$. It can be seen earlier that the introduction of hole convolution in the model structure can arbitrarily expand the receptive field without increasing the number of parameters, and at the same time keep the resolution of the feature map unchanged. The enlargement of the receptive field is conducive to the detection of larger defects, and the unchanged resolution of the feature map makes the feature map have more detailed features, which is conducive to the realization of the detection task in this paper. Therefore, through atrous convolution, the model has both a larger receptive field and a higher resolution and at the same time reduces the deviation caused by the multiple up-sampling operations of feature fusion, which can achieve a better detection effect. The specific design details of the optimized network structure are as follows:

1. The down sampling rate of the three stages from the fourth stage to the sixth stage is 16, which means that the feature map size of these three stages is $1/16$ of the original image size, but the size of the feature map in the fifth stage of the original residual network is $1/32$ of the size of the original image.

2. Two kinds of hole bottleneck structures are proposed, and the receptive field can be increased by using hole convolution. However, considering the amount of calculation and memory, the fifth and sixth stages have the same number of feature map channels.

3.5. Optimization of Pooling Algorithm. The core idea of RoI pooling is to use the nearest neighbor difference algorithm to realize the feature pooling process. The specific principle is shown in Figure 4. Taking the scarring defect map of a mechanical design product as an example, it is assumed that the RoI size of the scarring defect in the figure is $332 \times 332$. Since the down-sampling rate is 16, the corresponding size of defect RoI feature pooling is $332/16 = 20.75$. At this time, RoI pooling performs the first quantization, and the size of the feature map of the defective RoI is rounded to $20 \times 20$, as shown in Step 3 in Figure 4. Then the $20 \times 20$ area is fixed into a $7 \times 7$ feature submap. Since 20 divided by 7 is not an integer, RoI pooling is used for the second quantization, and the number is rounded to 2, starting from the upper left corner with 2 is the step size, and takes the maximum value in the $2 \times 2$ region as the output to obtain a $7 \times 7$ feature map.

It can be seen from Steps 4 and 5 in Figure 4 that after the complete defect feature in the original image is quantized and rounded twice by RoI pooling, the actual size of the defect feature detected in the fully connected layer of the R-CNN module is finally $224 \times 224$. However, the size of the defect feature in the original image is $332 \times 332$, which shows that the RoI quantization and rounding operation brings a large deviation. The calculation deviation caused by RoI pooling quantization rounding directly leads to the deviation of the pixels in the RoI area and the deviation of the corresponding position in space. Therefore, in essence, faster R-CNN does not achieve complete translation invariance, that is, when tracking from the convolutional mapping of RPN to the pixel mapping of the actual image, the structures of the two are different. The input and output of faster R-CNN are not pixel-to-pixel aligned, especially for small-sized defects such as mechanical design products, small deviations will also have a greater impact on the detection accuracy, including defect classification accuracy, and defect location regression accuracy.

4. Experiment and Analysis

Take a sample set of mechanical design product defects. In subsection 2, model the experiment after feature extraction algorithm optimization. Then, pooling algorithm optimization experiment and analyze it completely. Finally, combine all experiments and analyze them completely.
4.1. Establishment of a Sample Set of Mechanical Design Product Defects. Training a good deep learning model usually requires preparing a large amount of sample data. Generally speaking, as the amount of sample data increases, the generalization ability of the model will be improved, so in the case of a large number of sample images, the model has strong detection performance. A total of 3648 images with defects were collected in this paper, including 1224 scratches, 1312 scars and 1112 rust. In order to balance the sensitivity to different types of defects in the model training process, it is necessary to standardize the sample set uniformly, that is, select 1000 better images from each type of defect image as the original sample set, and discard the redundant images. After screening, each type of defect in the defect sample set has the same number of images, with a total of 3000 images for the three types of defects. Since the defect samples collected in this paper are not particularly large, the allocation ratio needs to be reasonably considered when dividing the sample set. After research and analysis, in this paper, the ratio of training set and test set is 4:1 to divide the sample data set for each defect, and randomly select 800 pieces of each kind of defect, a total of 2400 pieces are used as training samples, and the remaining 600 pieces are used as test pieces sample. In the data set of this paper, each defect image contains at least one defect, and some images contain multiple defects of different scales to ensure that the trained detection model can adapt to the detection of multiple types of defects at different scales.

The training of the model in this paper adopts the method of transfer learning, that is, pretraining on the public large data set and then training on the data set of this paper. The advantages are as follows: because the training is carried out on the basis of the pretrained model, so the training time is greatly shortened, and the results are generally ideal. When the data set is small, it can also train the ideal effect. After the faster R-CNN model was pretrained, it was trained on the data set of this paper for 100 epochs. In the model training process, TensorBoard is used to visualize the training process, so as to monitor the change of loss value in real time. Among them, the ordinate is the loss value, and the abscissa is the epoch. Figures 5 and 6 show the overall change of the loss of the two models during the training process. As the number of training increases, the loss gradually decreases and eventually stabilizes.

It can be seen from Figures 5 and 6 that the initial loss value of the faster R-CNN model is lower than that of the R-CNN model, the initial loss value of the R-CNN model is larger, and the model convergence speed is slower. In contrast, the faster R-CNN model has a faster convergence speed, the convergence effect is better.

4.2. Model Experiment after Feature Extraction Algorithm Optimization. After completing 100 training epochs, for the model saved after training, use the test set in the defect sample set to test the effect of the model and take the model with the best model test effect in this experiment as the final network model. Calculate the average precision (AP) value of the evaluation index of the model on each type of defect. As shown in Table 1, it is the AP value of the calculated three kinds of defects on the test set.
It can be seen from Table 1 that the AP value of the model optimized by the feature extraction algorithm for scratch defects is increased by nearly 3%, and the other two defects are also increased by more than 2%. It can be seen from the test results that the AP value of the faster R-CNN model optimized by the algorithm has been improved, and it can be seen that the detection model optimized by the feature extraction algorithm has a better detection effect.

4.3. Pooling Algorithm Optimization Experiment and Analysis. After completing 100 training epochs, for the model saved after training, the test set in the sample set is used to test the effect of the model, and the model with the best model test effect in this experiment is taken as the final network model. Calculate the AP value of the evaluation index of the model on each type of defect. As shown in Table 2, it is the AP value of the calculated three kinds of defects on the test set.

It can be seen from Table 2 that the AP value of the model optimized by the pooling algorithm for scale defects has increased by more than 5%, and the improvement of scar defects by more than 3%. In Table 2, the AP value of the improved model for each type of defect is averaged, and the mAP using the improved detection method is calculated to increase from 89.38% to 93.61%. It can be seen that the mAP value of the original faster R-CNN model is improved by 4.23% after the pooling algorithm optimization. It can be seen from the test results that the AP value of the faster R-CNN model after the algorithm optimization has been improved. As in the DD of mechanical design products, the detection precision of defects is an important index to evaluate the DD model. Therefore, comparing this detection index, it can be seen that the detection model optimized by the pooling algorithm has a better detection effect.

4.4. Combined Experiments and Analysis. The aforementioned two optimization methods optimize different parts of the original faster R-CNN, respectively, and it can be seen from the results of grouping experiments that the comprehensive detection performance of the model has been improved to a certain extent. Therefore, in the experiments in this section, the original faster R-CNN model was optimized by two optimization methods at the same time, and the experiment was performed again, and four experiments were compared with the two grouping optimization experiments and the original faster R-CNN experiment. The experimental results are shown in Table 3 which are the AP values of the three defects on the test set.

It can be seen from Table 3 that the AP value of the model optimized by the combined algorithm for scale defects has increased by more than 6%, scratch defects by more than 5%, and scar defects by more than 4%. In Table 3, the improved model of the combined algorithm averages the AP value of each type of defect and calculates that the mAP of the improved detection method using the combined algorithm is increased from 89.38% to 94.85%. It can be seen that the mAP value of the original faster R-CNN model is improved by 5.47% after the combined algorithm optimization. It can be seen from the test results that the faster R-CNN model optimized by the combined algorithm has the highest AP value compared with the other three groups of experiments. As in the detection of mechanical product design defects, the AP is the most important. Therefore, comparing the improved accuracy of the aforementioned algorithms, it can be seen that the detection model optimized by the combined algorithm has a better detection effect.

5. Conclusion

In the production process, it is difficult to prevent some mechanically designed products with surface defects and surface defects not only affect their appearance quality but also affect their actual performance. Traditional manual detection is inefficient and inaccurate; traditional image processing methods are easily disturbed by ambient light and background; the background of surface defects of
mechanically designed products is complex, and it is difficult to extract features manually; therefore, deep learning is introduced for workpiece surface DD and carry out related technical research. This paper collects three kinds of mechanical design product surface defects and analyzes the characteristics of the defects. A sample set of surface defects of mechanical design products based on deep learning is established. Research and propose to use the superior faster R-CNN model for experimental verification of the sample set in this paper. By analyzing the structure of the faster R-CNN model, it is proposed that the model adopts the advantages of residual structure and feature fusion in faster R-CNN, and it is optimized from two parts of the FEN and RoI pooling structure. For the shortcomings of the faster R-CNN model using the 13-layer convolution of VGG16 to extract features, a pattern of residual structure, feature fusion and hole convolution is proposed for optimization, which can increase the number of network layers, fuse the feature information between different layers, expand the receptive field, and improve the resolution of the feature map, so as to realize the detection of multiscale defects in the multilayer feature map. For the optimization method of the aforementioned two parts, two groups of experiments and one group of combination experiments were carried out, totaling three groups of experiments. It can be seen from the experimental results that both the individual optimization and the combined optimization of the two parts improve the detection performance of the model, and the average accuracy of the three kinds of defects after the optimization of the combined algorithm reaches 94.85%.

Data Availability
The data sets used during this study are available from the corresponding author on reasonable request.

Conflicts of Interest
The authors declare no conflicts of interest.

References
[1] Y. Huang, C. Qiu, and K. Yuan, “Surface defect saliency of magnetic tile,” The Visual Computer, vol. 36, no. 1, pp. 85–96, 2020.
[2] Y. Q. Wu and H. W. Deng, “Research on bearing surface DD based on support vector machine,” China Plant Engineering, vol. 413, no. 2, pp. 83–84, 2019.
[3] Z. Hong, B. Gao, Y. T. Gui, W. L. Woo, and L. Bai, “Metal defects sizing and detection under thick coating using microwave NDT,” Ndt & E International, vol. 60, pp. 52–61, 2013.
[4] X. Tao, D. Zhang, W. Ma, X. Liu, and D. Xu, “Automatic metallic surface defect detection and recognition with convolutional neural networks,” Applied Sciences, vol. 8, no. 9, p. 1575, 2018.
[5] H. Lin, B. Li, X. Wang, Y. Shu, and S. Niu, “Automated defect inspection of LED chip using deep convolutional neural network,” Journal of Intelligent Manufacturing, vol. 30, no. 6, pp. 2525–2534, 2019.
[6] M. Bentoumi, P. Aknin, and G. Bloch, “On-line rail defect diagnosis with differential eddy current probes and specific detection processing,” The European Physical Journal - Applied Physics, vol. 23, no. 3, pp. 227–233, 2003.
[7] X. Li, B. Gao, W. L. Woo, G. Y. Tian, X. Qiu, and L. Gu, “Quantitative surface crack evaluation based on eddy current pulsed thermography,” IEEE Sensors Journal, vol. 17, no. 2, pp. 412–421, 2016.
[8] B. Ji, Y. Li, D. Cao, C. Li, S. Mumtaz, and D. Wang, “Secrecy performance analysis of UAV assisted relay transmission for cognitive network with energy harvesting,” IEEE Transactions on Vehicular Technology, vol. 69, no. 7, pp. 7404–7415, 2020.
[9] X. Lin, J. Wu, S. Mumtaz, S. Garg, J. Li, and M. Guizani, “Blockchain-based on-demand computing resource trading in IoV-assisted smart city,” IEEE Transactions on Emerging Topics in Computing, vol. 9, no. 3, pp. 1373–1385, 2020.
[10] X. Xu, L. Huang, R. Wu et al., “Multi-feature fusion method for identifying carotid artery vulnerable plaque,” IRBM, vol. 9, 2021.
[11] Y. Li, W. Zhao, and J. Pan, “Deformable patterned fabric defect detection with Fisher criterion-based deep learning,” IEEE Transactions on Automation Science and Engineering, vol. 14, no. 2, pp. 1256–1264, 2016.
[12] J. Tian, J. Gao, and Y. Ao, “Automatic surface defect detection for mobile phone screen glass based on machine vision,” Applied Soft Computing, vol. 52, pp. 348–358, 2017.
[13] E. A. Smirnov, D. M. Timoshenko, and S. N. Andrianov, “Comparison of regularization methods for ImageNet classification with deep convolutional neural networks,” Aasri Procedia, vol. 6, pp. 89–94, 2014.
[14] M. D. Zeiler and R. Fergus, “Visualizing and Understanding Convolutional networks,” in Proceedings of the European Conference on Computer Vision, pp. 818–833, Springer, Cham, Zurich, Switzerland, September 2014.
[15] W. Yang, Q. Liu, S. Wang et al., “Down image recognition based on deep convolutional neural network,” Agricultural Information Processing, vol. 5, no. 2, p. 7, 2018.
[16] C. Szegedy, W. Liu, Y. Jia et al., “Going deeper with convolutions,” in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9, Boston, MA, USA, June 2015.
[17] S. Ioffe and C. Szegedy, “Batch normalization: accelerating deep network training by reducing internal covariate shift,” in Proceedings of the International Conference on Machine Learning, pp. 448–456, PMLR, Lille, France, July 2015.
[18] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818–2826, San Juan, PR, USA, June 2016.
[19] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778, Las Vegas, NV, USA, June 2016.
[20] H. Zheng, L. X. Kong, and S. Nahavandi, “Automatic in-car tire surface defect detection processing,” The European Physical Journal - Applied Physics, vol. 29, no. 2, pp. 412–421, 2016.
[21] Y. Wang and J. Su, “Automated defect and contaminant inspection of HVAC duct,” Automation in Construction, vol. 41, pp. 15–24, 2014.
[22] Y. F. Zhao, C. Gao, and J. G. Wang, “Research on machine vision based industrial product surface defect detection algorithm,” Computer Applications and Software, vol. 29, no. 2, pp. 152–154, 2012.
[23] M. Guo, L. Hu, and J. Zhao, "Surface defect detection method of ceramic bowl based on Kirsch and Canny operator," *Acta Optica Sinica*, vol. 36, no. 9, Article ID 0904001, 2016.

[24] J. Masci, U. Meier, G. Fricout, and J. Schmidhuber, "Multi-scale Pyramidal Pooling Network for Generic Steel Defect classification," in *Proceedings of the The 2013 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, Dallas, TX, USA, August 2013.

[25] V. Natarajan, T. Y. Hung, S. Vaikundam, and L. T. Chia, "Convolutional networks for voting-based anomaly classification in metal surface inspection," in *Proceedings of the IEEE International Conference on Industrial Technology (ICIT)*, pp. 986–991, IEEE, Toronto, Canada, March 2017.

[26] T. Wang, Y. Chen, M. Qiao, and H. Snoussi, "A fast and robust convolutional neural network-based defect detection model in product quality control," *International Journal of Advanced Manufacturing Technology*, vol. 94, no. 9, pp. 3465–3471, 2018.

[27] J. Chen, Z. Liu, H. Wang, A. Nunez, and Z. Han, "Automatic defect detection of fasteners on the catenary support device using deep convolutional neural network," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 2, pp. 257–269, 2018.

[28] R. Ren, T. Hung, and K. C. Tan, "A generic deep-learning-based approach for automated surface inspection," *IEEE Transactions on Cybernetics*, vol. 48, no. 3, pp. 929–940, 2017.