

Research Article

Casting Defect Detection and Classification of Convolutional Neural Network Based on Recursive Attention Model

Zhichao Zhao and Tiefeng Wu 

School of Information and Control Engineering, Qingdao University of Technology, Qingdao 266520, Shandong, China

Correspondence should be addressed to Tiefeng Wu; wutiefeng@qut.edu.cn

Received 17 August 2022; Revised 20 September 2022; Accepted 27 September 2022; Published 14 October 2022

Academic Editor: M. A. Rashid Sarkar

Copyright © 2022 Zhichao Zhao and Tiefeng Wu. 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 method based on deep learning shows excellent performance in the recognition and classification of surface defects of some industrial products. The method based on deep learning has high efficiency in the identification and classification of surface defects of industrial products, and the false detection rate and missed detection rate are relatively low. However, the recognition accuracy of defect detection and classification of most industrial products needs to be improved, especially for those with similar contours and relatively large structural different casting. This paper takes casting defect detection as the goal and proposes a convolutional neural network casting defect detection and classification (RCNN-DC) algorithm based on the recursive attention model. Through this model, the casting can be better identified and detected, and casting defects can be avoided as much as possible, which is of great significance to the technological development of the industry. First, use a large amount of readily available defect-free sample data to detect anomalous defects. Next, we compare the accuracy and performance of the detection model and the general recognition model. The research results show that the test effect of the RCNN-DC casting defect detection network model is significantly better than the traditional detection model, with a classification accuracy of 96.67%. Then, we compare the RCNN-DC network with three classic popular networks, GooGleNet, ResNet-50, and AlexNet. Among them, AlexNet and ResNet-50 achieved 95.00% and 95.56% classification accuracy, respectively, while GooGleNet achieved slightly better results of 96.38%. In contrast, the accuracy of RCNN-DC is 1.67% higher than that of AlexNet, while the number of FLOPs is reduced by 17.2 times, and the accuracy is 1.09% higher than that of ResNet-50, while the number of FLOPs is reduced by 99.7 times, and the accuracy is higher than GooGleNet 0.29% while FLOPs whose number has been reduced by 36.5 times.

1. Introduction

The classification of casting surface defects is a key link in industrial defect detection. However, in traditional industries, this link is often performed manually. Artificial castings are prone to stomata, sand sticking, sand inclusion, trachoma, expanding sand, cold insulation, and insufficient pouring. In order to replace manual operation, people hope that the machine can use computer vision technology to automatically detect the surface defects of the casting. Because the image of casting surface defects is affected by light and material changes, and there are large differences in appearance of defects within castings, and defects between classes have similar aspects, it is still a huge challenge to use computer vision technology to classify defects. Attention

region extraction contrast experiment is to verify image enhancement, image filtering, interference suppression, and so on.

With the development of machine vision technology, some optical inspection methods based on machine vision are introduced to the defect detection of castings. Among them, the hybrid defect detection method for semiconductor castings proposed by REN R applies image alignment strategy and adaptive image difference method to the detection of surface bumps on castings [1]. He et al. proposed a point pattern matching positioning algorithm based on azimuth environment characteristics to realize the detection of casting pin defects [2]. Ren uses Faster R-CNN to detect the defects of the capping surface of the oil pepper filling production line [3]. Liu proposed that machine vision and

image processing technology have been continuously developed and researched and applied more and more widely. Similar defect detection research mainly includes binary threshold segmentation, feature extraction and selection, and morphological processing [4]. Wang binarizes the model and uses morphological filtering to separate defects, and finally performs logical classification by calculating gradient histogram features [5].

At present, there have been research studies on the application of lightweight networks in casting defects and classification. For example, Sadykova D proposed a MagnetNets network for real-time detection of surface defects of ceramic tiles [6]. Fuqi proposed an improved MobileNet network for classifying liver pathological tissues [7]. Ma Peng combined gray-level co-occurrence matrix and density clustering to analyze the surface defects of brake pads and then extract features based on the texture of the brake pads for detection [8]. Buslaev processes multiple images separately and merges them into one image, extracts features from it and the previous processing steps, establishes the feature space of the learning classifier based on artificial neural network, and performs automatic detection and classification [9]. Prappacher has proposed a fast and reproducible convolutional neural network for detecting surface defects in roller bearing rolling elements. In order to solve the problem of high requirements for the neural network method to mark training data, the use of simulated defects makes the detection system faster and more cost-effective than existing systems [10].

This paper takes casting defect detection as the goal and proposes a convolutional neural network casting defect detection and classification algorithm based on the recursive attention model. The common recognition model compares accuracy and performance. The castings enter their respective sorting tanks according to the classification results. The system process ends to complete the detection and classification of the bottom face defects of the gasket. In this paper, the detection of castings mainly adopts the SA layer module under the recursive attention model, and the recognition adopts the method of convolutional neural network. After the bottom face detection is completed, the qualified castings are turned over and tested again through the inspection system. The system process ends to complete the top and bottom face defects of the gasket.

2. Recursive Attention Model and Convolutional Neural Network Casting Defect Detection and Classification Method

2.1. Advantages of Recursive Attention Model and Convolutional Neural Network in Casting Defect Detection and Classification. Traditional image processing and neural network training methods are mainly used for defect detection at home and abroad. The neural network requires a large number of training samples, but it is very difficult to obtain defect samples, so the operation is more difficult [11]. Therefore, in order to improve the accuracy and efficiency of the detection of casting defects, and realize the automatic

detection of gaskets, according to the characteristics of complex background and many interference of gasket image, the collected image is analyzed and processed, and the defects of casting are detected and classified, and a set of defect detection and classification system of casting end face is designed and developed. After experiments, the system is stable and the detection results meet the requirements. One type is a two-stage classification algorithm based on region recommendation. Representative algorithms include R-CNN, Fast R-CNN, and Faster R-CNN. The other is an end-to-end one-stage regression algorithm. Representative algorithms include SSD, YOLO, YOLOV2, and YOLOV3. Among them, the two-stage classification algorithm has higher detection accuracy and the single-stage regression algorithm has faster detection speed [12].

YOLOV3 draws on the residual network (ResidualNetwork) structure to form a deeper network level, and on the premise of maintaining the speed advantage of the single-stage regression algorithm, it improves the detection accuracy and the detection ability of small targets. Aiming at the problem of small defect targets in waveguide castings and large scale changes between different types of defects, the original YOLOV3 algorithm feature extraction network, multiscale feature fusion, and a priori target frame design are improved [13]. In the production process of RCNN-DC, due to the influence of aging of the manufacturing process and equipment, there will be defects such as chipping and damage to the positioning column. In addition, the production of RCNN-DC has very strict environmental requirements. Due to the effects of temperature and manufacturing environment, casting in the waveguide region will be defective. The defect size of RCNN-DC varies greatly, the target of the casting in the waveguide area is small, and the maximum size of the chipping reaches (33×282), while the minimum size of the waveguide casting is (19×23). This is the difficulty of RCNN-DC defect detection. In order to strengthen the detection accuracy of small targets, YOLOV3 adopts an idea similar to FPN to perform feature fusion on the three-scale feature maps obtained by 32, 16, and 8 times downsampling of the input image through the backbone network, thereby achieving multiscale prediction [14]. In this paper, the attention extraction model of ROI is mainly used, and the accuracy of defect detection is improved within a certain range through the SA layer recognition module. DBL (Darknetconv2d_BN_Leaky) is the basic component of YOLOV3, which consists of a convolutional layer (Conv), a batch normalization layer (BN), and an activation layer (LeakyRelu) [15]. Resn is a large component of YOLOV3, including n Res_units that introduce a residual structure. Res_unit draws on the idea of ResNet and performs a residual layer jump connection after two DBLs. The backbone network Darknet-53 of YOLOV3 uses five Resn structures Res1, Res2, Res8, Res8, and Res4 [16]. In the prediction stage, for each input image, we predict a three-dimensional tensor including the position of the target bounding box, the target category, and the confidence level [17]. Then, we divide the 13×13 , 26×26 , and 52×52 feature maps into $S \times S$ cells (S is 13, 26, and 52, respectively), and set 3 a priori boxes for each grid to predict. There are 3 bounding boxes, and the output

dimension is $S \times S \times (3 \times (4 + 1 + M))$, which is the offset of 4 bounding boxes, 1 defect target confidence level, and M defect types [18].

2.2. Recognition Method of Recursive Attention Model for Castings. The Recursive Attention Module (SALayer) is a multiscale context cascading module, including a top module and a multilevel pooling module. The top module directly uses 1×1 convolution processing on the input features to obtain spatial global features [19]. The multilevel pooling module contains 3 downsampling parts to obtain three different resolution features of $1/2$ input image, $1/4$ input image, and $1/8$ input image, respectively, and obtain context information of different scales until the input is restored feature size [20]. Finally, it is combined with the top module level as the final output of the structure [21]. Through multiple downsampling, the features of castings of different sizes are obtained, and convolution layers are used for feature selection and extraction. Both large areas can be used to obtain high-level context information, and small areas can also be used to obtain low-level context information. This combination of levels can maintain the hierarchical dependencies of different scales, adapt to feature mapping at different scales, and avoid other the misclassification of symbiotic and mixed castings in land [22]. The expression of recursive attention is shown in formula (1):

$$\ln\left(\frac{FI_{it}}{FI_{it}-1}\right) = \alpha + \beta \ln FI_{it} - 1 + \phi X_{it} - 1 + \nu_i + \tau_t, \quad (1)$$

$$E = \frac{\sum_{j=1}^k \sum_{h=1}^k \sum_{i=1}^{n_j} \sum_{r=1}^{n_h} |y_{ji} - y_{hr}|}{2n^2 u}.$$

Compared with natural images, the coverage of casting images is very small; there are large differences in the coverage area of casting types, and the situation of coexistence and mixing with castings and many other casting types. Natural images have a large projection range and relatively low cost, the image is realistic, and can also face relatively complex scenes. Some casting types such as circuit lines may have a small area in the image, while some casting types such as screws occupy a relatively small area. Different casting types have different spaces [23]. Taking into account the influence of the casting space coverage and the surrounding casting types, a recursive attention module based on multilayer scales is designed. The structure simulation formula is as follows:

$$u_h \leq u_j \leq \dots \leq u_k,$$

$$E_j = \frac{(1/2u_j) \sum_{i=1}^{n_j} \sum_{r=1}^{n_j} |y_{ji} - y_{jr}|}{n_j^2}, \quad (2)$$

$$Ew = \sum_{j=1}^k G_{jj} p_j s_j.$$

This module can well integrate multiple spatial features of different scales [24]. After transformation, another form of circle equation can be obtained as follows:

$$E_{jh} = \frac{\sum_{Z=1}^{h_j} \sum_{r=1}^{n_h} |y_{ji} - y_{hr}|}{n_j n_h (u_j + u_h)},$$

$$E_{nb} = \sum_{j=2}^k \sum_{h=1}^{j-1} G_{jh} (p_j s_h + p_h s_j) D_{jh}, \quad (3)$$

$$E_t = \sum_{j=2}^k \sum_{h=1}^{j-1} G_{jh} (p_j s_h + p_h s_j) D_{jh} (1 - D_{jh}).$$

Only the parameter E_t is needed to get the center and radius of the circle. Let the coordinates of the point on the subpixel contour be (j, h) ; D_{jh} is the sum of squares of the casting model:

$$D_{jh} = \frac{d_{jh} - p_{jh}}{d_{jh} + p_{jh}}. \quad (4)$$

Affected by the casting material, the gray value of the washer end face is evenly distributed, and the gray value of the defect is large, but the difference with the surface is small, so it is difficult to separate the defect. If the gray value distribution of the end face of the gasket is not uniform, the gray value will be small, which is not conducive to comparison, and the difference of the gasket is difficult to distinguish. Image multiplication is used to significantly increase the difference between bright and dark image points in order to improve contrast. According to the input image, the gray value range is obtained, and the gray adaptive factor Mult is multiplied to transform the image formula as follows:

$$d_{jh} = \int_0^\infty dF_j(y) \int_0^y (y-x) dF_h(x), \quad (5)$$

$$d_{jh} = \int_0^\infty dF_h(y) \int_0^y (y-x) dF_j(y) * \text{Mult}.$$

With the increasing demand of running high-quality deep neural network on embedded devices, the research on lightweight network model design is encouraged [25]. These networks tend to use “sparse connection” convolution, which can reduce the computational cost and hinder the information exchange between groups. Based on ShuffleNet and MENet, and mixconv convolution module, this paper proposes a RCNN-DC network. While the classification accuracy of the network is improved, the calculation cost is also reduced. Channel shuffle is the core of the shuffle net and aims to solve the problem that group convolution interferes with the exchange of information between groups and leads to poor performance. When the number of channels in each group is 3, channel shuffling cannot completely avoid the loss of information between groups. Each group in the second volume layer only receives one channel from each group in the first volume layer, resulting in the other two channels in each group being ignored. Therefore, most of the inter group information cannot be

used. This problem is more serious in more channel groups. With the increase of the number of groups, the number of channels in each group increases. However, the number of channels received by the second convolution layer is still one, and the number of ignored channels in each group also increases, resulting in serious information loss between groups and a significant decline in network performance. Hybrid convolution module deep convolution (DWConv) is becoming more and more popular in modern lightweight networks. Hybrid convolution module deep convolution can be used to classify images or process signals in reality, and can be used on image or tensor input. The commonly used deep convolution divides each channel into a group for group convolution, which greatly reduces the amount of parameters and the cost of calculation. However, the traditional method is to simply use 3×3 convolution kernel, ignoring the size of convolution kernel. Referring to Mix-Conv's idea of multicore combination, using convolution cores of different sizes to replace deep convolution, the large convolution core can improve the accuracy of the model in a certain range, and the multiconvolution core can improve the fitness of the model in different resolutions.

2.3. Extraction Model of Attention Area (ROI). In order to solve the problem of information loss between groups, this paper uses the ideas of merge and evolution in MENet for reference. A narrow feature map is used to fuse and encode the channel information between groups, and the matching transform is combined with the original network to obtain more distinguishing features. The purpose of fusion is to aggregate all channel features and encode the information between groups to form a narrow feature map. Based on the original feature map generated by group convolution, the network carries out fusion coding transformation to achieve the purpose of aggregating the features of all channels. Spatial transformations allow you to extract more spatial information, normalize batch processing, and activate ReLU without changing the number of channels. In order to combine the processed feature graph with the original network to obtain more distinguishing features, the network carries on the matching transformation to the spatial feature graph and then carries on the batch processing normalization and sigmoid activation to obtain the matching feature graph with the same dimension as the original feature graph. In order to further improve the matching ability of neural network elements, the product of neural network elements and neural network features is combined. In the process of channel fusion, the information of each channel is encoded by 1×1 single point convolution operation, so in the final matching feature map, the transformed channel contains the information from each channel of the original feature map, which avoids the loss of information between groups in the convolution process.

In the process of image processing, the scope of image processing needs to be limited to a specific attention area in the image, namely ROI. In the casting image, the attention area is the middle ring area; in addition to the fixture and background interference, these interference areas need to be

suppressed or eliminated. Because the average gray values of background, fixture, and gasket are close, it is difficult to extract ROI by using the commonly used binarization and connected area. In this paper, we need to filter the Fourier frequency domain by selecting a low-pass filter, filter out some small texture and point noise, and eliminate interference. Therefore, it is necessary to use the characteristics of ROI as a ring and use the least square method to fit the inner and outer diameter circle to generate the region and find its difference set. Based on this, an algorithm for extracting the ring ROI is proposed, which is aimed at the extraction of the attention region and is aimed at the gray background and difficult to separate. Taking the contour of the outer circle as an example, the least square method is used to approximate the circle:

$$f(x) = \frac{1}{Nh} \sum_{i=1}^N k\left(\frac{X_i - x}{h}\right), \quad (6)$$

$$k(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right),$$

where $f(x)$ is a digital image of size $m \times n$. The spectrum diagram (7) obtained by Fourier transform is as follows:

$$h_t = \tanh(w_c x_t + u_c (r_t \Theta h_{t-1}) + b_c), \quad (7)$$

$$h_t = z_t \Theta h_{t-1} + (1 - z_t) \Theta h_t. \quad (8)$$

In order to facilitate filtering and spectrum analysis, the original spectrum diagram is centralized to obtain the centralized spectrum diagram, as shown in (8), which contains three kinds of parameter amplitude, phase, and energy spectrum describing characteristic information:

$$P = \sigma t, \quad (9)$$

$$= \frac{\sqrt{(1/n) \sum_{i=1}^n (FI_{it} - FI_{it})^2}}{FI_{it}}.$$

The traditional method of feature extraction algorithm plus classifier design is a mainstream surface defect classification method in the current industry. By adopting this as a benchmark experiment and comparing it with the RCNN-DC network, the reliability of the experiment can be ensured. In the process of image acquisition, the environment, illumination, and other uncertain factors will more or less affect the imaging, and the noise will be unevenly distributed in the image:

$$u_{(j|i)} = w_{ij} A_i, \quad (10)$$

$$s_j = \sum_i c_{ij} u_{(j|i)}.$$

The centralized spectrum of Fourier is distributed in the center of divergence, and there are many discrete high-frequency points. Therefore, the low-pass filter is selected to filter the Fourier frequency domain, filter out some small texture and point noise, and eliminate interference. Part of the contour information of defects will be lost, but it has

limited influence on the extraction of defects. The low-pass filter function is as follows:

$$\ln\left(\frac{PI_{it}}{PI_{it}-1}\right) = \alpha + \beta \ln PI_{it} - 1 + v_i + \mathfrak{F}_r,$$

$$k_{t1}[i] = \sum_j \cos(w_i^1, w_j^2), \quad (11)$$

$$\theta = -\frac{1}{T} \ln(1 + \beta).$$

After the above image enhancement and image filtering, the original gray image is clipped, and the definition domain of the image is reduced to the specified target area, that is, the ring area of the above attention. The interference of other backgrounds and fixtures is eliminated, and the detection of the end ring is carried out, which greatly reduces the detection time and improves the efficiency and accuracy.

3. Convolution Neural Network Casting Defect Detection and Classification Based on Recursive Attention Model

3.1. Model. This paper takes casting defect detection as the goal and proposes a neural network-based casting defect detection algorithm (RCNN-DC), which uses only simple and a large amount of nondefect sample data to detect abnormal defects. The algorithm proposed in this paper includes two stages: the neural network training stage of the brain neural model and the surface area detection and classification stage. The neural network training stage of the brain nerve model is mainly to train the casting surface model and determine its surface accuracy. Surface area detection mainly uses algorithms to classify castings that do not meet the conditions.

3.2. Steps. (1). Environment and Training

The algorithm is implemented on PyTorch framework. The experimental environment is Intel Xeon w3175x processor, 512 gb running memory, NVIDIA Quadro M4000 graphics card, and the operating system is arch Linux. In the experiment, 10,000 images are randomly selected from the dataset as the training set. Before training, the edge of the data image is adjusted to 256 pixels, and then the center is cut to 224×224 pixels. There are other interference areas, and the attention area is closed by rectangular structure elements. The defect area is contained in the attention circle area, so the image needs to be thresholded. The difficulty of threshold segmentation is how to find the threshold accurately when isolating defects. The selection of threshold directly determines the accuracy of region segmentation and the accuracy of subsequent region feature analysis. Combining the advantages and disadvantages of each method and the actual detection situation, the maximum interclass variance method, also known as Ostu segmentation method, is selected to obtain the best threshold T , and the image is segmented into the front and background. If the pixel value is greater than the threshold T , it is the attention target; if it is

less than the threshold T , it is the background area. After the closed operation, some small interference points are removed, and the attention area is smooth and complete. Then, the connected area is disconnected to divide the attention area into corresponding independent areas, and the separate areas are processed separately to achieve the purpose of separating defects.

(2) The specific process of the system is as follows:

Step 1. the casting is transported by the conveyor belt, the conveyor belt is driven by the stepper motor, and the photoelectric sensor triggers the timing. After reaching the image acquisition area, the conveyor belt stops, and the fixture starts to fix the casting. After the fixture is in place, the PLC sends out the image acquisition signal to trigger the camera acquisition.

Step 2. after the image acquisition is completed, the computer sends the acquisition completion signal to PLC, drives the fixture to reset, and the conveyor belt continues to work.

Step 3. after arriving at the classification station, the image is processed by computer to complete the defect detection, and the corresponding photoelectric sensor is triggered for classification.

Step 4. according to the classification results, the castings enter the respective sorting slot, and the system process ends to complete the detection and classification of the defects on the lower end face of the washer. After the detection of the lower end face, the qualified castings are turned over and tested again through the detection system, and the system process ends to complete the detection and classification of the defects on the upper and lower end faces of the washer. In this way, the upper end face can be detected for defects, and the area of the surface can be kept flat, which is also conducive to the classification of castings.

4. Convolution Neural Network-Based Defect Detection

4.1. Accuracy Analysis of Casting Defect Detection Model. The image data collected by the image acquisition system is shown in Figure 1. It can be seen from the figure that the attention area is circle, but there are interferences such as fixtures and background in the image, and the surface texture is also highlighted when the defect is highlighted by illumination. Aiming at defect detection in this complicated environment, we proposed an image processing algorithm based on Fourier transform for detecting cast end face defects. Due to the similarity of the end faces and limited space, the following end faces are taken as an example for research, and the upper-end faces are tested using the same algorithm.

The recognition of casting is shown in Table 1. Firstly, the color image is collected and transformed into gray image. The inner and outer diameter circle area is obtained by binarization. The subpixel edge is extracted to obtain the RCNN-DC contour. The subpixel circle contour is fitted by

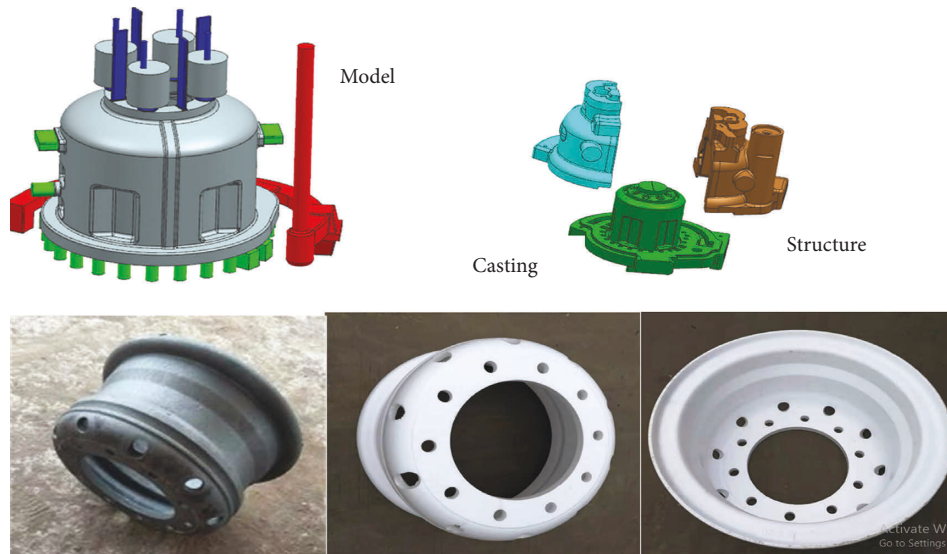


FIGURE 1: Image acquisition system collects casting image.

TABLE 1: The color image collected by the camera converted to grayscale.

| Item | Ostu | XLD | Defect detection | Image processing | PyTorch | PLC |
|------------|------|------|------------------|------------------|---------|------|
| Fixture | 1.46 | 3.02 | 1.94 | 3.36 | 2.97 | 2.88 |
| Attention | 4.71 | 5.21 | 3.34 | 2.63 | 5.11 | 5.15 |
| Background | 5.46 | 3.46 | 5.13 | 3.14 | 5.1 | 3.77 |
| RCNN-DC | 2.56 | 3.98 | 4.77 | 3.22 | 1.92 | 2.8 |
| Fourier | 5.1 | 4.21 | 4.98 | 4.08 | 3.44 | 3.21 |
| Casting | 2.88 | 6.61 | 3.32 | 3.05 | 5.2 | 2.71 |

the least square method to extract the attention circle area. At the same time, the filter suppresses other interference to highlight the defect area. Finally, the processed gray image is cut; through the gray histogram threshold segmentation, the defect area can be obtained quickly, and the defect area can be classified by analyzing the characteristics of the defect area.

As shown in Figure 2, the parameters reflect the parameters of the model, and the accuracy reflects the accuracy of the model in the defect classification task. The average running time is 12 consecutive runs of a single image. After removing the maximum and minimum values, take the average result of 10 runs to measure the speed at which the model runs. In order to test the comprehensive performance of the network model in the task of casting surface defect classification, this paper introduces four evaluation indexes: flops, parameters, precision, and average running time, and uses PyTorch-OpCounter tool to test the flops and parameters of the network. Among them, flops refer to floating-point operations, which is used to measure the complexity of the model. The lower the complexity is, the lighter the model is. In general, the flops of lightweight network can be reduced to below 150×10^6 .

As shown in Table 2, RCNN-DC is compared with some of the most advanced network structures and benchmark experiments on the NEU-CLS dataset. MFLOPS in the table is a unit used to measure flops, which means 106 floating-point operations per second.

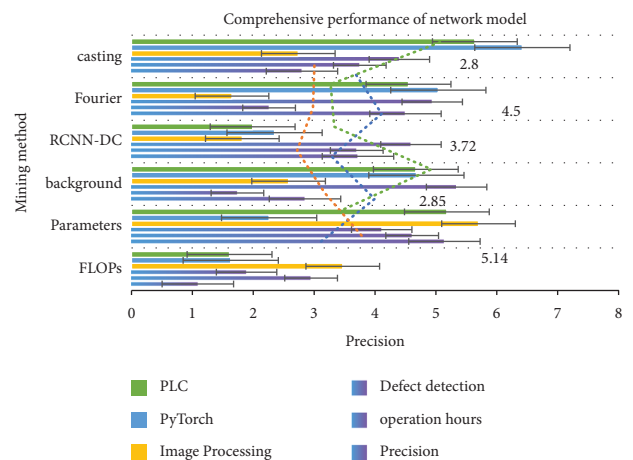


FIGURE 2: Comprehensive performance of network model.

As shown in Figure 3, with the addition of the hybrid convolution module, the model recognition accuracy is improved, and RCNN-DC achieves the best accuracy of 98.61%, which indicates that the hybrid convolution can effectively reduce the sensitivity of the network to the large convolution core, improve the network stability, and optimize the model performance. In order to make a comprehensive comparative experiment, this paper introduces SENet with channel attention mechanism for comparison.

As shown in Table 3, although the classification accuracy of SE-ResNet-50 and SE-ResNet-101 is higher than that of

TABLE 2: RCNN-DC and some of the most advanced network structures.

| Item | Precision | Operation hours | Defect detection | Image processing | PyTorch | PLC |
|------------|-----------|-----------------|------------------|------------------|---------|-----|
| FLOPs | 53 | 58 | 59 | 62 | 56 | 53 |
| Parameters | 75 | 79 | 67 | 74 | 79 | 65 |
| Background | 83 | 96 | 90 | 87 | 87 | 83 |
| RCNN-DC | 98 | 92 | 90 | 80 | 85 | 96 |
| Fourier | 85 | 93 | 93 | 98 | 95 | 94 |
| Casting | 99 | 84 | 100 | 98 | 80 | 88 |

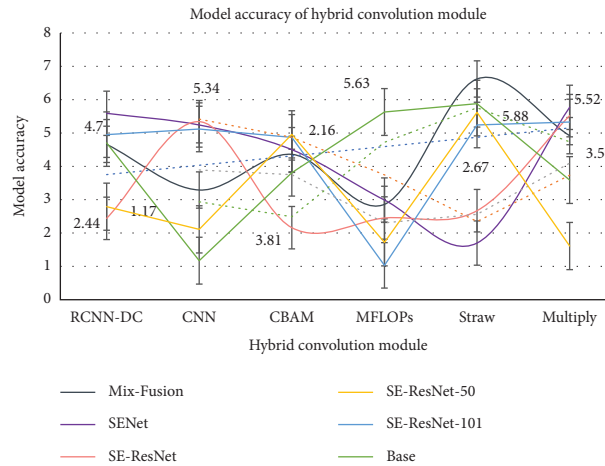


FIGURE 3: Model accuracy of hybrid convolution module.

TABLE 3: The classification accuracy of SE-ResNet-50 and SE-ResNet-101.

| Item | Mix-fusion | SENet | SE-ResNet | SE-ResNet-50 | SE-ResNet-101 | Base |
|----------|------------|-------|-----------|--------------|---------------|------|
| RCNN-DC | 4.66 | 5.59 | 2.44 | 2.79 | 4.95 | 4.7 |
| CNN | 3.29 | 5.24 | 5.34 | 2.11 | 5.12 | 1.17 |
| CBAM | 4.36 | 4.5 | 2.16 | 4.96 | 4.87 | 3.81 |
| MFLOPs | 2.86 | 2.99 | 2.45 | 1.72 | 1.03 | 5.63 |
| Straw | 6.62 | 1.7 | 2.67 | 5.62 | 5.24 | 5.88 |
| Multiply | 4.92 | 5.77 | 5.52 | 1.61 | 5.33 | 3.59 |

RCNN-DC (35) and RCNN-DC (3579), the recognition accuracy of RCNN-DC is still slightly better. At the same time, the operation speed of RCNN-DC network is much faster than that of SE-ResNet network, and the parameter quantity and complexity of SE-ResNet network are much larger than those of mix fusion network. RCNN-DC model uses SSIM + L1 combined loss function, which can be used for defect detection of regular and irregular surface texture at the same time. For regular texture samples, only using L1 loss, that is, the weight coefficient $\alpha = 1$, can get better detection results; for irregular texture samples, different weight coefficients will produce different detection results. This experiment uses irregular texture samples in the graph, the weight coefficient α ranges from 0 to 1, and the step size is set to 0.1, which is used to adjust the proportion of SSIM loss and L1 loss.

This paper analyzes the convergence and stability of attention model convolutional neural network. The training loss curve and verification accuracy curve of RCNN-DC, mix fusion, ShuffleNetV2, and MobileNetV2 are shown in Figure 4. The RCNN-DC network has the fastest convergence

speed, the loss is stable at 0.004, the ShuffleNetV2 has the second highest convergence speed, the loss is stable at 0.005, the MobileNetV2 has the slowest convergence speed, and the loss is stable at 0.009. In the figure, the verification accuracy of mix fusion is significantly higher than that of the other two networks, and the trend of accuracy curve shows that the stability of mix fusion and ShuffleNetV2 is higher than that of MobileNetV2. In order to ensure the effectiveness of the model, this paper further compares the comprehensive performance of RCNN-DC with three advanced lightweight networks, ShuffleNetV2, and MobileNetV2.

As shown in Table 4, in order to verify the impact of the proposed different modules on the network performance, this paper designs a RCNN-DC (base) network which only uses traditional deep convolution. The basic network structure is consistent with the casting defect detection network model, and only the mixed convolution in the module diagram is replaced by the 3×3 deep convolution with the same convolution kernel scale. The results show that the casting defect detection network model is better than the traditional detection model, and the classification accuracy is

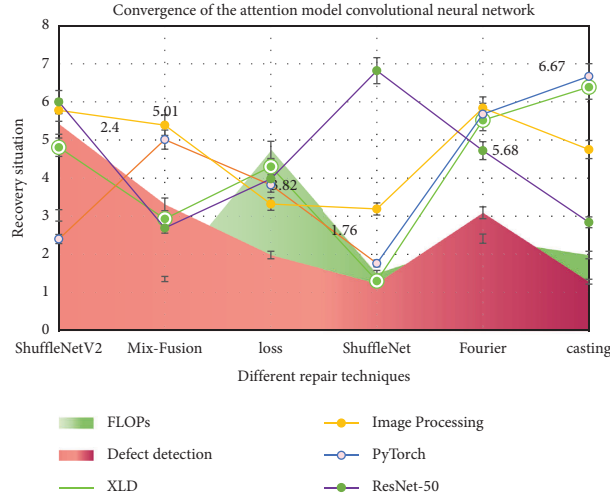


FIGURE 4: Convergence of the attention model convolutional neural network.

TABLE 4: The impact of different modules on network performance.

| Item | FLOPs | XLD | Defect detection | Image processing | PyTorch | ResNet-50 |
|--------------|-------|------|------------------|------------------|---------|-----------|
| ShuffleNetV2 | 3.02 | 4.81 | 5.42 | 5.78 | 2.4 | 6 |
| Mix-fusion | 1.35 | 2.93 | 3.31 | 5.39 | 5.01 | 2.69 |
| Loss | 4.73 | 4.3 | 1.98 | 3.32 | 3.82 | 3.98 |
| ShuffleNet | 1.5 | 1.3 | 1.24 | 3.19 | 1.76 | 6.82 |
| Fourier | 2.41 | 5.52 | 3.09 | 5.84 | 5.68 | 4.72 |
| Casting | 1.98 | 6.39 | 1.28 | 4.75 | 6.67 | 2.84 |

96.67%. After that, we compare the three popular networks, i.e. DC GINet and GooxNet. Among them, AlexNet and ResNet-50 achieved 95.00% and 95.56% classification accuracy, respectively, while GoogleNet achieved a slightly better result of 96.38%. In contrast, when RCNN-DC (base) is 1.67% higher than AlexNet, the number of flops is reduced by 17.2 times, when RCNN-DC (base) is 1.09% higher than ResNet, the number of flops is reduced by 99.7 times, and when RCNN-DC (base) is 0.29% higher than GoogleNet, the number of flops is reduced by 36.5 times.

The experimental results of casting defect samples under various loss functions are shown in Figure 5. In this paper, the defect image samples with irregular surface texture are used. From the comparison of residual results, it can be seen that MSE is used as the loss function of the algorithm to get the residual results. In addition to the real defect area, there are more noise points in other areas, forming pseudodefects; while using SSIM as the loss function, the detected defect area is slightly smaller than the real defect area; compared with other loss functions, the structure loss function is better the combination of loss function SSIM and loss function L1 achieves good results. In this paper, the defect image samples with regular surface texture are used. From the comparison of residual results, it can be seen that the integrity of defect area obtained by MSE loss function is poor, which is similar to the detection result of MSE + SSIM combination. The combination of the structural loss function SSIM and the L1 loss function achieves good results, and the detection results are similar to using only the L1 loss function.

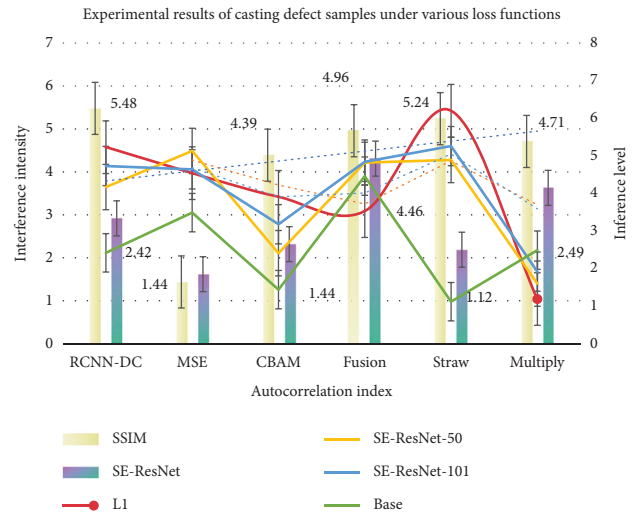


FIGURE 5: Experimental results of casting defect samples under various loss functions.

As shown in Table 5, for the defect samples with irregular surface texture in the table, the combination of structural loss function SSIM and RCNN-DC-11 loss function achieves better results in recall rate and RCNN-DC, which is slightly inferior to SSIM in accuracy rate; for the defect samples with regular surface texture in the table, only RCNN-DC loss function achieves the highest recall rate, and the combination of SSIM and L1 loss function is the best For F1 measure, the RCNN-DC loss function performs best.

TABLE 5: The combination of the structural loss function SSIM and the L1 loss function.

| Item | SSIM | L1 | SE-ResNet | SE-ResNet-50 | SE-ResNet-101 | Base |
|----------|------|------|-----------|--------------|---------------|------|
| RCNN-DC | 5.48 | 5.23 | 2.92 | 4.17 | 4.73 | 2.42 |
| MSE | 1.44 | 4.54 | 1.62 | 5.13 | 4.64 | 3.49 |
| CBAM | 4.39 | 3.91 | 2.32 | 2.41 | 3.18 | 1.44 |
| Fusion | 4.96 | 3.53 | 4.31 | 4.82 | 4.84 | 4.46 |
| Straw | 5.24 | 6.2 | 2.19 | 4.89 | 5.26 | 1.12 |
| Multiply | 4.71 | 1.19 | 3.63 | 1.6 | 1.92 | 2.49 |

4.2. *Casting Defect Classification Effect Based on Convolution Neural Network with Recursive Attention Model.* As shown in Figure 6, RCNN-DC outperforms ShuffleNet in classification performance. Considering that the number of channels with reduced branches in RCNN-DC is less than that in ShuffleNet, we attribute this improvement to the proposed fusion coding operation. Although ShuffleNet has more channels, it still suffers from intergroup information loss, and RCNN-DC makes effective use of intergroup information. Therefore, RCNN-DC has more distinguishing features than ShuffleNet and overcomes the problem of performance degradation. At the same time, compared with ShuffleNetV2 and MobileNetV2, the classification accuracy of RCNN-DC is 1.36% and 1.67% higher than that of ShuffleNetV2 and MobileNetV2, respectively.

The dependence analysis of RCNN-DC network on data is shown in Table 6. The proportion of training data represents the percentage of the new training set in the original training set. When the proportion of training data is 50% and 25%, the network accuracy decreases slightly; when the proportion of training data is 10% or less, the network performance will greatly decline, and there is a slight over fitting phenomenon, which indicates that the network has a certain dependence on the amount of data.

The results of various defect detections are shown in Figure 7. The defect detection and performance analysis experiments verify the effectiveness of the proposed defect detection method and image processing algorithm. The images collected by the actual casting detection system are applied to the proposed image processing method. Combining the strengths and weaknesses of each method with the actual detection situation, the maximum class variance method, also known as the Ostu segmentation method, is selected to obtain the optimal threshold T and split the image into front and background. If the pixel value is greater than the threshold T , it is the attention target; if it is less than the threshold T , it is the background area.

In order to verify the accuracy of casting defect extraction, detection experiments were carried out for three common types of defects. Six samples of castings were selected for experiments and the detection results were recorded. The defect detection results are shown in Figure 8. Therefore, RCNN-DC can achieve better classification accuracy and faster running speed with less computational cost in comparison with traditional texture feature extraction methods, several classic networks, or advanced lightweight networks. On the one hand, the

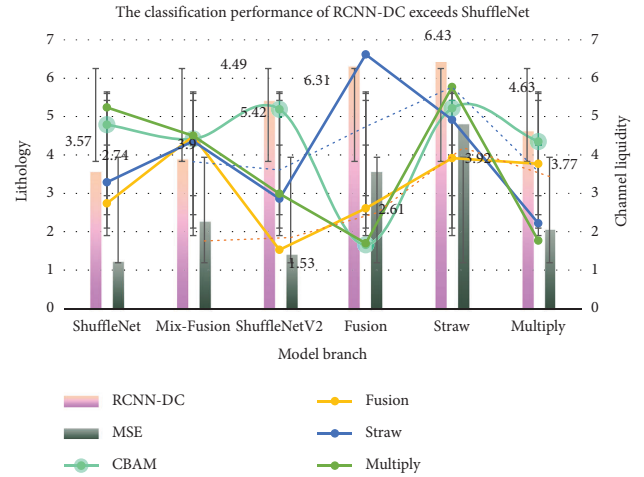


FIGURE 6: The classification performance of RCNN-DC exceeds ShuffleNet.

TABLE 6: Dependence of the RCNN-DC network on data.

| Item | RCNN-DC | MSE | CBAM | Fusion | Straw | Multiply |
|--------------|---------|------|------|--------|-------|----------|
| ShuffleNet | 3.57 | 1.24 | 4.79 | 2.74 | 3.29 | 5.24 |
| Mix-fusion | 3.9 | 2.28 | 4.42 | 4.49 | 4.36 | 4.5 |
| ShuffleNetV2 | 5.42 | 1.42 | 5.19 | 1.53 | 2.86 | 2.99 |
| Fusion | 6.31 | 3.57 | 1.66 | 2.61 | 6.62 | 1.7 |
| Straw | 6.43 | 4.81 | 5.22 | 3.92 | 4.92 | 5.77 |
| Multiply | 4.63 | 2.07 | 4.35 | 3.77 | 2.22 | 1.77 |

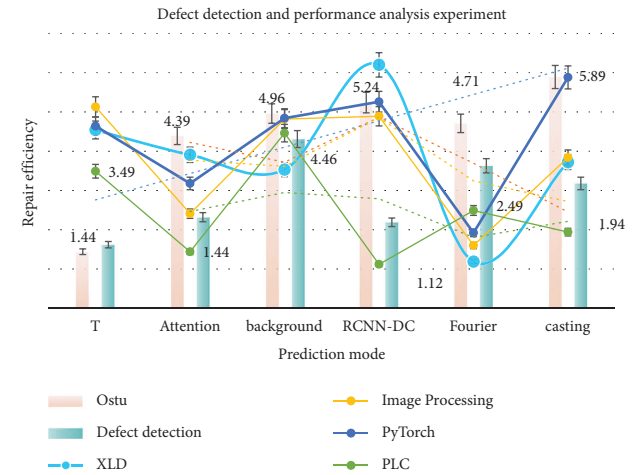


FIGURE 7: Defect detection and performance analysis experiment.

fusion branch of MF module effectively solves the problem that “sparse connection” convolution hinders the information exchange between groups, and the reduced parameter branch greatly reduces the amount of calculation of the network; on the other hand, compared with the traditional deep convolution, the hybrid convolution reduces the sensitivity of the network to large convolution core and improves the stability of the network.

In this paper, the direct traditional binary threshold segmentation method and a series of improved image preprocessing and thresholding segmentation are used

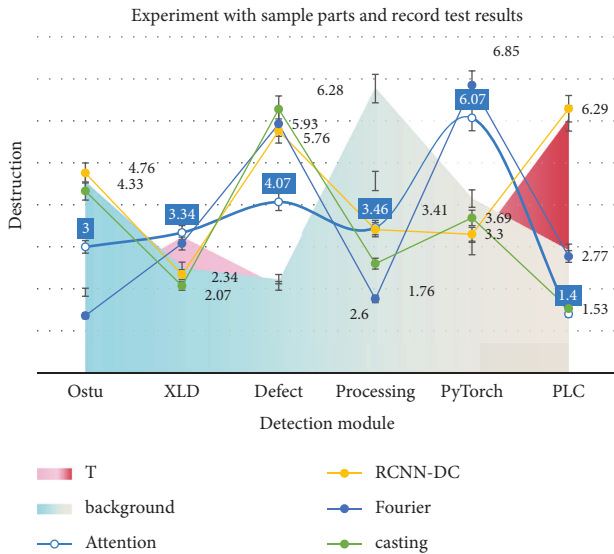


FIGURE 8: Experiment with sample parts and record test results.

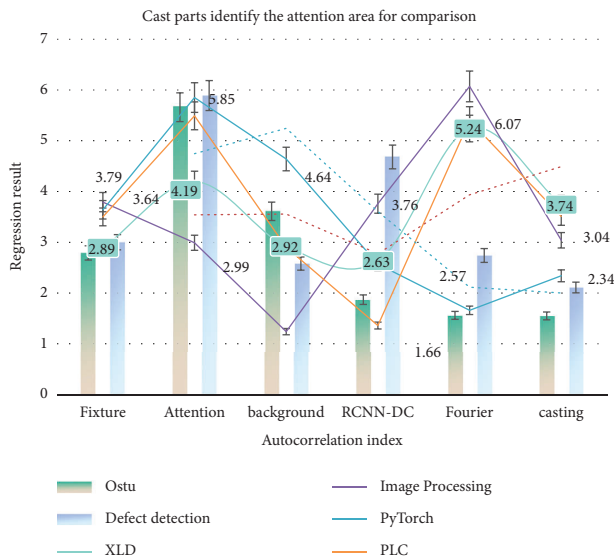


FIGURE 9: Cast parts identify the attention area for comparison.

respectively to compare the attention regions obtained by the two methods, and the effect is shown in Figure 9. For the original image, using the traditional binarization and threshold segmentation method, the number of connected components is 10379, which is difficult to separate the defect area through the subsequent multithreshold and multifeature screening. After improving and using a series of image processing, the number of regional connected components is greatly reduced to 105. Through the Fourier transform low-pass filtering, most of the texture and point noise are eliminated. Although there are still interference areas, most of them are large area and circular area. Through the subsequent shape selection and other features, the defect area can be quickly and accurately segmented so as to extract the image attention deficit areas.

5. Conclusions

In this paper, a convolution neural network casting defect detection and classification algorithm based on recursive attention model is proposed, which only uses a large number of defect free sample data to detect abnormal defects. The algorithm proposed in this paper includes two stages: convolution neural network training stage of recursive attention model and surface defect detection and classification stage. RCNN-DC casting defect detection network model test effect is significantly better than the traditional detection model, classification accuracy of 96.67%. Then, RCNN-DC network is compared with three classic popular networks, GooGleNet, ResNet-50, and AlexNet. Among them, AlexNet and ResNet-50 achieved 95.00% and 95.56% classification accuracy, respectively, while GooGleNet achieved a slightly better result of 96.38%. In contrast, when RCNN-DC is 1.67% higher than that of AlexNet, the number of flops is reduced by 17.2 times, when RCNN-DC is 1.09% higher than that of ResNet-50, the number of flops is reduced by 99.7 times, and when RCNN-DC is 0.29% higher than that of GooGleNet, the number of flops is reduced by 36.5 times.

In order to solve the problems of convolutional neural network, such as the increasing number of parameters, the increasing cost of computation, and the existing lightweight network cannot completely avoid the loss of information between groups, this paper proposes a novel lightweight network RCNN-DC based on ShuffleNet and MEnet. Experimental results show that networks avoid loss of information between groups, improve classification accuracy, reduce computational costs and parameters, improve comprehensive performance compared to other networks, and cast surface defect classification tasks. Strong support is shown for deployment on mobile devices. The next work of this paper is to further verify the generalization performance of the model and carry out the optimization algorithm, adaptation equipment, and other work to meet the needs of production.

The software and hardware of the casting detection system are designed and discussed to realize the purpose of casting defect detection and classification. In order to improve the efficiency of the detection and the complexity of the detection process, an automatic detection system is designed according to the characteristics of castings and defects to realize image acquisition and defect recognition and classification. Taking the lower end face of casting as the detection object, the upper end face and the lower end face adopt the same method. According to the characteristics of casting and image, aiming at the noise interference, defect classification, and other problems, firstly, the image blob analysis is used to find out the area where the inner and outer diameter circle is located, and the subpixel contour is fitted with the least square method to get its size parameters, and the difference set of the inner and outer diameter circle is generated, Fourier transform, low-pass filter convolution, and inverse Fourier transform are used to highlight the defects and suppress the interference of other factors. Finally, the defects are obtained and the

features are extracted, and the defects are classified according to the features so as to achieve the purpose of defect detection and classification. After the above image enhancement and image filtering, the original gray image is clipped, and the definition domain of the image is reduced to the specified target area, that is, the ring area of the above attention. The interference of other backgrounds and fixtures is eliminated, and the detection of the end ring is carried out, which greatly reduces the detection time and improves the efficiency and accuracy.

Data Availability

The data underlying the results presented in the study are available within the manuscript.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by Cooperative Education Project of the Ministry of Education (202102413008, 202102413009).

References

- [1] 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, 2018.
- [2] Y. He, K. Song, Q. Meng, and Y. Yan, "An end-to-end steel surface defect detection approach via fusing multiple hierarchical features," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 4, pp. 1493–1504, 2020.
- [3] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.
- [4] Z. Liu, X. Miao, J. Chen, and H. Jiang, "Review of visible image intelligent processing for transmission line inspection," *Power System Technology*, vol. 44, no. 3, pp. 1057–1069, 2020.
- [5] Y. Wang, Q. Li, and B. Chen, "Image classification towards transmission line fault detection via learning deep quality-aware fine-grained categorization," *Journal of Visual Communication and Image Representation*, vol. 64, no. 4, pp. 102647–103107, 2019.
- [6] D. Sadykova, D. Pernebayeva, M. Bagheri, and A. James, "IN-YOLO: real-time detection of outdoor high voltage insulators using UAV imaging," *IEEE Transactions on Power Delivery*, vol. 35, no. 3, pp. 1599–1601, 2020.
- [7] F. Ma, B. Wang, X. Dong, H. Wang, P. Luo, and Y. Zhou, "Power vision edge intelligence: power depth vision acceleration technology driven by edge computing," *Power System Technology*, vol. 44, no. 6, pp. 2020–2029, 2020.
- [8] P. Ma and Y. Fan, "Small sample smart substation power equipment component detection based on deep transfer learning," *Power System Technology*, vol. 44, no. 3, pp. 1148–1159, 2020.
- [9] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: fast and flexible image augmentations," *Information*, vol. 11, no. 2, pp. 125–215, 2020.
- [10] N. Prappacher, M. Bullmann, G. Bohn, F. Deinzer, and A. Linke, "Defect detection on rolling element surface scans using neural image segmentation," *Applied Sciences*, vol. 10, no. 9, pp. 3290–3340, 2020.
- [11] F. Hemmati, M. Miraskari, and M. S. Gadala, "Application of wavelet packet transform in roller bearing fault detection and life estimation," *Journal of Physics: Conference Series*, vol. 1074, no. 1, pp. 140–142, 2018.
- [12] P. Cawley, "The impedance method of non-destructive inspection," *NDT International*, vol. 17, no. 2, pp. 59–65, 1984.
- [13] S. J. Kim, "Damage detection in composite laminates using coin-tap method," *Journal of the Acoustical Society of America*, vol. 123, no. 5, pp. 3064–3068, 2008.
- [14] C. Yang and X. X. Yang, "Early fault diagnosis of rolling bearing based on GRD and TEO," *Journal of Vibration and Shock*, vol. 39, no. 13, pp. 224–229, 2020.
- [15] F. Lu and J. Q. Huang, "Feature extraction algorithm of clustering based on grey relational theory," *Systems Engineering-Theory & Practice*, vol. 32, no. 4, pp. 872–876, 2020.
- [16] S. G. Sun, Q. Zhang, and T. H. Du, "Fault diagnosis method of frame-type circuit breaker based on grey correlation degree," *Chinese Journal of Scientific Instrument*, vol. 38, no. 10, pp. 2525–2535, 2017.
- [17] D. Y. Wang, X. K. Wang, and W. W. Yu, "Off-Axis LED curved array lighting design for leather defect detection," *Laser & Optoelectronics Progress*, vol. 56, no. 8, pp. 182–202, 2019.
- [18] D. J. Li and R. H. Li, "Mug defect detection method based on improved faster RCNN," *Laser & Optoelectronics Progress*, vol. 57, no. 4, pp. 415–419, 2020.
- [19] G. S. Zhang, G. Y. Ge, and R. H. Zhu, "Gear defect detection based on the improved YOLOv3 network," *Laser & Optoelectronics Progress*, vol. 57, no. 12, pp. 121–129, 2020.
- [20] J. Liu, T. Shi, K. W. Wang, Z. Tang, and G. Liao, "Defect detection of flip-chip solder joints using modal analysis," *Microelectronics Reliability*, vol. 52, no. 12, pp. 3002–3010, 2012.
- [21] C. S. Chen, C. L. Huang, and C. W. Yeh, "A hybrid defect detection for in-tray semiconductor chip," *International Journal of Advanced Manufacturing Technology*, vol. 65, no. 1–4, pp. 43–56, 2013.
- [22] G. Liao, L. Du, L. Su, M. Zeng, L. Nie, and T. Shi, "Using RBF networks for detection and prediction of flip chip with missing bumps," *Microelectronics Reliability*, vol. 55, no. 12, pp. 2817–2825, 2015.
- [23] Z. Xu, T. Shi, X. Lu, and G. Liao, "Using active thermography for defects inspection of flip chip," *Microelectronics Reliability*, vol. 54, no. 4, pp. 808–815, 2014.
- [24] Z. S. Xu, T. L. Shi, X. N. Lu, and Q. Xia, "Failures detection of flip-chip using active thermography method based on wavelet transform," *Infrared and Laser Engineering*, no. 10, pp. 3233–3237, 2020.
- [25] L. Feng and Z. H. Gong, "An algorithm for chip surface defect detection based on sequential similarity and light source automatic adjustment," *Modern Electronics Technique*, vol. 40, no. 5, pp. 58–62, 2017.