IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 19 August 2022, accepted 18 October 2022, date of publication 26 October 2022, date of current version 3 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3217238

RESEARCH ARTICLE

Research on Improved Residual Network Classification Method for Defect Recognition of Thermal Battery

WENCHAO XU[1,2], (Member, IEEE), SIXIANG ZHANG[1], (Member, IEEE), FANG BAI[2], AND TAO ZHAO[1], (Member, IEEE)

[1]College of Mechanical Engineering, Hebei University of Technology, Tianjin 300131, China
[2]College of Information Engineering, Tianjin University of Commerce, Tianjin 300134, China

Corresponding authors: Sixiang Zhang (13502063552@163.com) and Wenchao Xu (xuwenchao@tjcu.edu.cn)

This work was supported by the Tianjin Education Commission Scientific Research Project: Research on the Development of Collaborative Innovation under the Background of the Integration of Production and Education in Colleges under Grant 2018SK094.

ABSTRACT Thermal battery is an ideal power supply for military applications such as artillery and ship equipment. Due to the sheet-type process of the thermal battery, various installation error defects occur in the assembly of thermal battery. Aiming at the problems of low efficiency and low defect-recognition rate of thermal battery detection, a thermal battery defect detection model is proposed based on residual network. First, the squeeze-and-excitation networks (SENet) structure based on the attention mechanism is introduced into residual block of the residual neural network, the connection between the feature extraction channels is established, and the improved deep residual network I-ResNet50 is obtained; Second, in order to prevent overfitting, the defect images processed in the production line and the laboratory are data-enhanced and labeled. Transfer learning strategy is introduced into the recognition model I-ResNet50, and then the training set data samples are input into the recognition model I-ResNet50 for training, and the activation function LReLu and Dropout skills are introduced to improve the classification ability of the I-ResNet50 model; Finally, the recognition model I-ResNet50 is applied to the test set and validation set, and each defect of the thermal battery are output. Comparison experiments are tested under different migration strategies and different optimizers and learning rates, and comparison experiments with the five classic network structures of ResNet50, YOLOV3, MobileNetV2, VGG16, and YOLOV4 are also tested. The test data show that the recognition accuracy rates of qualified images and the three types of defective images (Qualified Assembly, Missing Current Plate, Wrong Number of Stacks, and Reverse Stack) can reach 99.64%, 98.17%, 99.11%, and 95.40%, respectively, the overall recognition accuracy rate can reach 98.10%. The test results illustrate the model can detect thermal battery defects more accurately and quickly, and has good defect diagnosis ability, which is nearly 5% higher than the traditional method, and a new solution for defect detection in practical industrial scenarios of thermal battery is provided.

INDEX TERMS Transfer learning, residual network, squeeze-and-excitation networks, defect identification, thermal battery.

I. INTRODUCTION

Thermal battery is called thermal activated reserve battery. It uses molten salt as the electrolyte, and its heat source is ignited by an automatic activation mechanism and then melts and activates the electrolyte [1], [2]. It belongs to a storage type high-temperature molten salt primary battery, and both high-power discharge and low-power long-term discharge are possible. Therefore, the thermal battery has excellent performance, and it is very ideal as a military power supply, mostly applied for the power supply of artillery, ship equipment, and satellite navigation instruments [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Zhiyi Li.

Decades ago, the service life of the battery is greatly extended. However, with the use of this process, various installation errors have occurred in the assembly of the thermal battery, which requires various methods to be used for defect detection in the production line [4]. It is an efficient, time-saving, and labor-saving detection method to process the thermal battery imaging by X-ray machine to find defects.

Due to the difference in the internal density and thickness of thermal battery, the degree of X-ray absorption is also different. After the X-ray passes through the thermal battery, the intensity distribution is different. It contains the distribution information of the internal structure of the thermal battery. The X-rays carrying the information are converted into visible light distributions of different intensities through the scintillator, which is collected by the photoelectric sensor and transmitted to the computer through the acquisition circuit to generate thermal battery X-ray image.

The method of imaging and processing images with X-ray machines is widely used in various defect detection applications, such as the detection of solar photovoltaic panels, metal surfaces, wafers, various cracks, stains, etc. The traditional defect detection algorithm roughly includes the following steps: (1) Select an area. The whole image is traversed by the sliding window strategy, and the target information contained in the image is captured by changing the scale and proportion of the window. (2) Feature extraction. Visual information in candidate regions is extracted by hand-designed geometric feature templates. (3) Design a classifier and train the sample data. Finally, the detected redundant target frame is eliminated by post-processing, and the optimal position of the target object is found [5].

Defect detection methods based on traditional algorithms generally include methods based on image segmentation, feature extraction, and spectral algorithms.

A. IMAGE SEGMENTATION

In the defect detection method of image segmentation, the target and background areas of the image to be inspected are mainly segmented by the threshold, iteration, and edge calculation, so that defects are detected. And these different segmentation methods consist of threshold-based, region-based, and edge-based segmentation methods [6], [7].

The threshold-based segmentation method is to set a threshold to segment the image according to certain criteria. The pixel value is compared with this threshold, and those greater than or equal to the threshold are classified into one class, and those less than the threshold are classified into another class so that the target and the background are obtained. The most typical threshold segmentation method is the OTSU method [8], this method has a wide range of applications for defect detection due to its low computational complexity and is not easily affected by brightness and contrast. Literature [9] proposed the OTSU method, which used a parameter between 0 and 1 as the weight of the variance between classes, which is convenient to select the best threshold for segmentation. This method can obtain a good segmentation effect and detection rate in detecting surface defects.

The method based on region segmentation is to divide the image to be inspected into different regions according to certain features [10], [11]. The region segmentation method has the advantages of high robustness and is not easy to be disturbed by noise, so it is applied in the defect detection field. Reference [12] proposed an improved method to extract surface defects of silicon wafers, which accurately extracted the location of defects and the contours of closed regions.

The edge-based segmentation method has a good detection effect and fast speed for images with obvious boundaries and is favored in the field of defect detection. Reference [13] proposed an edge-based morphological segmentation algorithm to properly extract the surface defects of sewers, morphological features can be measured, and defects in sewer surfaces can be ineffectively diagnosed.

B. METHODS BASED ON FEATURE EXTRACTION

The method based on feature extraction mainly uses the feature information in the image, and these features are classified by the classifier to achieve defect detection. Features commonly used in defect detection often consist of corner features, texture features, and color features. Common detection methods based on corner features include boundary curvature-based detection method [14], Harris-based corner detection method [15], and SIFT-based corner detection method [16], etc. Common detection methods based on color features include the cumulative color histogram method, histogram intersection method [17], and center distance method [18]. Texture feature methods include the gray level co-occurrence matrix method [19] and the autocorrelation function method. Reference [20] proposed a detection algorithm to detect fabric defects using histogram features extracted from saliency maps, which has the ability to automatically detect defects in non-patterned fabrics and patterned fabrics. Reference [21] proposed an integrated defect detection method. The defect information on the tile surface is obtained by a gray level co-occurrence matrix, and then the surface defects are classified by a vector machine. This method has an excellent performance in both defect detection and eliminating the factors of non-conforming products in the ceramic tile industry.

C. METHODS BASED ON SPECTRAL ALGORITHMS

The frequency-domain features are mainly obtained by the spectral transformation in the method based on the spectral algorithm for the detection and classification of the defects. Commonly used spectral algorithms include Fourier transform method [22] and the wavelet transform-based method. Tsai et al. used the Fourier method based on image reconstruction for the detection of fabric surface defects, and this method is also suitable for the detection of other textured objects [23]. In [24], Li et al. proposed a wavelet transform-based surface defect detection method for solar flakes. The weight is set as the wavelet coefficient difference between

adjacent resolutions to detect whether there are defects on the solar surface.

Traditional defect detection methods are easily affected by subjective judgments and lead to unstable detection accuracy and low efficiency, which can no longer meet production needs. Traditional machine vision-based detection methods have achieved reliable results in many cases, but it requires specific preprocessing methods and specialized knowledge to extract representative features. In order to adapt to the development of the intelligent trend of modern manufacturing, more and more researchers have begun to study advanced detection methods. As a simple and fast data analysis method, deep learning has made breakthroughs in defect detection and other fields.

In the last few decades, deep learning has made outstanding contributions to defect detection and classification problems. Among the many branches of deep learning, CNN is one of the common frameworks. In 1998, LeCun [25] improved the LeNet structure which is applied to digit recognition, and achieved excellent results. Alex Krizhevsk et al. proposed the AlexNet network model, which won the first prize in the ImageNet image classification competition in 2012. The model achieved breakthrough improvements in speed and accuracy. CNN is widely concerned by a large number of scholars for its excellent feature extraction ability and generalization ability, and it has great potential in various fields.

The development of classification network models has moved in the direction of various deep network structures. In 2013, ZFNet [26] is proposed and achieved good classification results, a new feature visualization technology, including de-convolution, de-pooling, deactivation function, etc., is introduced into the model, and this technology is used for deep analysis and understanding of structure. The VGGNet proposed by Simonyan et al. in 2014, which formed a 16 to 19 deep network structure by stacking 3×3 convolution kernels and 2×2 pooling layers, it is further improved the classification accuracy by learning larger spatial features, and GoogleNet and Inception are introduced for the first time. Multiple convolutional layers and pooling layers are packaged into a unit, and the generalization performance is improved by setting a 1×1 convolution kernel. For the VGGNet network, a smaller convolution kernel is used for dimensionality reduction operation in GoogleNet, and it has a deeper network structure, but VGGNet has a more prominent classification performance in single-network [27]. 50-layer ResNet network not only retains the complete picture feature information through shortcuts and skips connections, but also the gradient disappearance and gradient explosion phenomenon generated by the deep network during training are alleviated. With the continuous development of classification networks, more and more classification networks are applied to the task of surface defect classification. Reference [28] proposed a five-layer network structure method for supervised steel defect classification, this method achieved excellent results with an error rate of only 7% in 7 categories of defect classification tasks collected from the actual production line. Reference [29] used ReLU as the activation function and analyzed image data for specific problems, and realized the classification of track surface defects. Research [30] introduced a model for classifying, a single CNN model is adopted by the model, and effective defect features are extracted. Thus, the automatic classification of various types of wafer surface defects is realized.

Many universities, research institutes, and leading enterprises did great research in defect detection field and proposed many excellent target detection models, which are suitable for surface defects detection of objects. Girshick first used R-CNN in 2014 [31]. R-CNN is a region-based convolutional neural network. It first uses the selective search method to extract some candidate regions that may have targets from the input image, and then extracts the features of each candidate region through AlexNet, and uses SVM to determine whether there is a target in the candidate region. The algorithm achieved good detection and recognition results, its accuracy reached 66% on the VOC2007 dataset. However, due to a large amount of computational allowance and tedious training steps, the algorithm takes about 47s to process a picture, which cannot be put into practical use. With the advent of Fast-CNN and Faster R-CNN [32], detection models are moving towards higher detection accuracy while reducing the training rate. Redmon et al. proposed YOLO [33] in 2016, which abandoned the step of obtaining candidate regions in the R-CNN series, and output bounding boxes by dividing the feature map into multiple cells. It has the advantages of fast speed and strong generalization ability. However, YOLO is insufficient, for shortcomings of YOLO method Ma et al. proposed the SSD model in [34], combined with YOLO and the anchor box idea of Faster R-CNN, 6 feature maps of different scales are extracted through the VGG-16 basic network. The fully connected layer is removed so that the detection accuracy is improved.

However, to strengthen the ability of SSD for small object detection, DSSD [35] and FSSD [36] also appeared. These network models have their advantages and are widely used in defect detection tasks. Reference [37] proposed a novel vision-based method, which combined SSD and YOLO frameworks to form a deep CNN model, which can well identify surface defects of high-speed rail firmware. Kang [38] et al. proposed an improved SSD method, which can effectively classify the surface defects of hot-dip galvanized sheets, and the detection accuracy and efficiency are improved.

The traditional machine vision method has great limitations and cannot be changed according to the type of the recognized target, resulting in a high false-positive rate of manually designed template matching. Therefore, the neural network algorithm does not need to re-develop the model when dealing with new defect types, and the maintenance cost is low. More importantly, compared with traditional defect detection algorithms, deep learning models have unique advantages such as strong algorithm adaptability, transferable learning in similar situations, and experience reuse [39].

Transfer learning and the improved SE convolution module are introduced and applied to the thermal battery image recognition network. The main contributions are summarized below:

(1) The SENet structure is introduced into the residual block of the residual neural network to establish the connection between the feature extraction channels, and the improved deep residual network model I-ResNet50 is obtained;
(2) The defect images are data-enhanced and labeled. Transfer learning strategy is introduced into the recognition model I-ResNet50, and the activation function LReLu and Dropout skills are introduced to enhance the classification ability of the I-ResNet50model;
(3) Under different migration strategies and different optimizers and learning rates, comparison experiments with the five classic network structures of ResNet50, YOLOV3, MobileNetV2, VGG16, and YOLOV4 are tested. The recognition accuracy rates of qualified images and the three types of defective images (Qualified Assembly, Missing Current Plate, Wrong Number of Stacks, Reverse Stack) can reach 99.64%, 98.17%, 99.11%, and 95.40%, respectively, the overall accuracy can reach 98.10%, it achieved a major performance improvement of 5% over traditional methods proposed by Zhao[40].

Section 2 represents augmentation of thermal battery defect image, deep residual network, and transfer learning. Section 3 introduces the establishment of I-ResNet50 and Adam optimizer, SGD optimizer, and batch normalization in detail. Section 4 represents four groups of comparative experimental analysis, visualization of the feature map, analysis of recognition rates, and confusion matrix. Some concluding remarks are drawn in Section 5.

II. MATERIALS AND METHODS

A. MATERIALS AND EQUIPMENT

Since CNN is used to identify various defects of the thermal battery, its learning effect mainly lies in the scale and diversity of samples. However, there is no public image library for thermal batteries, and the defect images of different types of thermal batteries can only be preprocessed manually. And the convolutional neural network may have an overfitting problem due to insufficient data. Overfitting makes the generalization ability of the model worse, and cannot be well used in the detection of new data sets. To prevent over-fitting, the dataset images in this study are enhanced by the image data enhancement module in Keras. These operations include translation, rotation, noise addition, etc. The data processing operations are as follows:

(1) Image translation.
(2) The image is flipped up and down, left and right.
(3) Zoom: The image zoom ratio is 1.0-1.2, and the reduction ratio is 0.8-1.0.

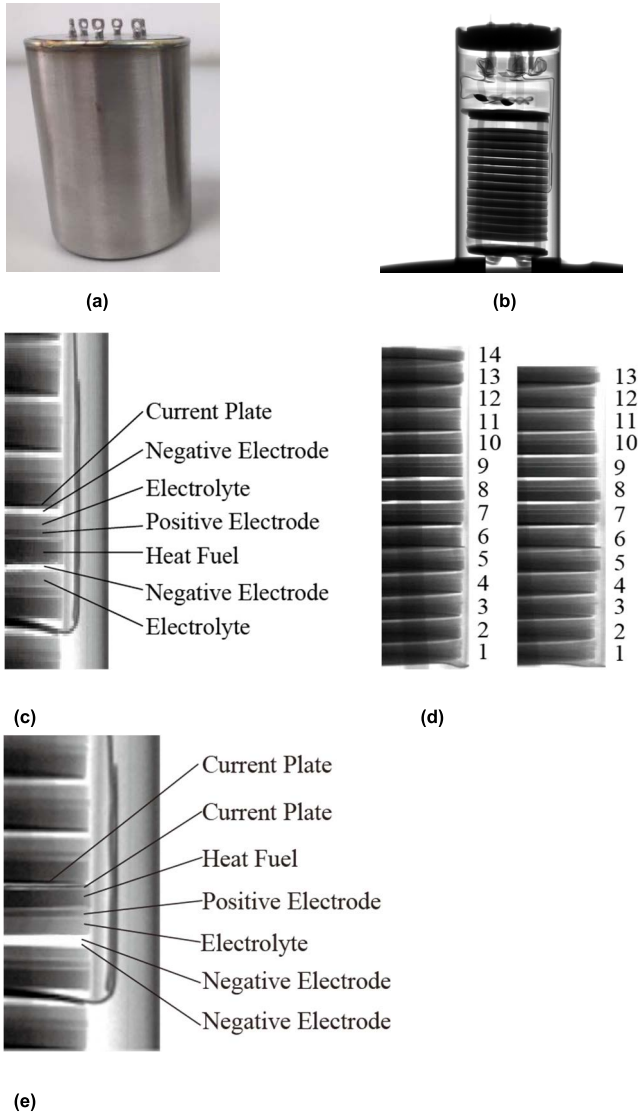


FIGURE 1. The 3 categories of defective and qualified images of thermal battery images.

(4) Rotation: Rotate 0°-30° counterclockwise or clockwise.
(5) Add Gaussian noise and salt and pepper noise, etc.
(6) Increase or decrease brightness: adjust the light intensity of the X-ray machine according to the imaging situation.
(7) Increase contrast: adjust the ratio to 0.9-1.1.
(8) Image mirroring.
(9) Perform histogram homogenization for some underexposed images to improve the noise problem caused by light.

Effective image data augmentation can further improve classification accuracy. A total of 1560 thermal battery images with qualified assembly and three categories of defects are randomly selected to form the sample dataset, and which are resized to 224 × 224 pixels in batches. In view of

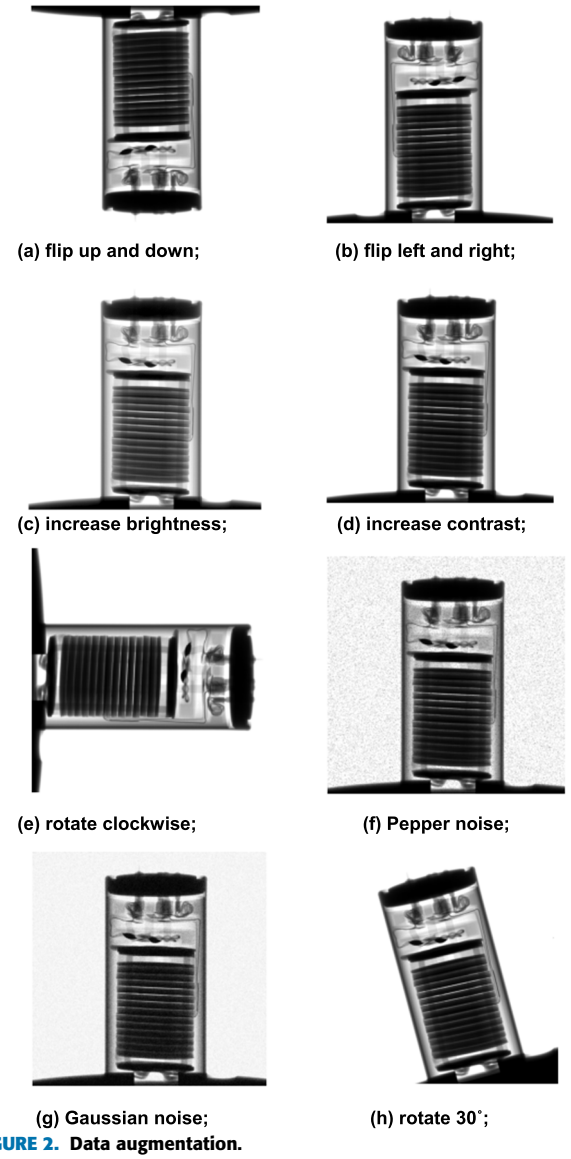


FIGURE 2. Data augmentation.

the fact that the thermal battery will not change its characteristics when observed from different angles, a total of nine transformation methods are used to expand the sample data set, such as random scaling, mirroring, random brightness adjustment, and Gaussian noise, etc. A total of 10,015 thermal battery images are obtained, including Qualified Assembly, Missing Current Plate, Wrong Number of Stacks(The correct number of stacks inside the thermal battery is 14, the wrong number is 13.), Reverse Stack defect images, which greatly increased the diversity of samples. The training set, test set, and validation set are divided according to the ratio of 6:2:2. The 3 categories of defective images and qualified images of the thermal battery assembly error defect dataset are shown in Figure 1. Some image samples after data enhancement are shown in Figure 2.

The test platform consists of computer hardware and a development platform. Computer Hardware: The processor CPU is AMD R5 3600X @ 4.25GHz; The Graphics Processing Unit (GPU) is an NVIDIA GeForce RTX2060. Development platform: Pytorch deep learning framework on Windows 10 operating system configured with NVIDIA CUDA Toolkit 10.1 and deep neural network acceleration library NVIDIA CUDNN v8. 0. 4. The programming language is Python3.7.4.

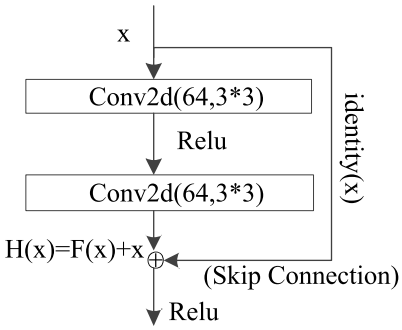


FIGURE 3. Block of residual learning.

B. DEEP RESIDUAL NETWORK

The emergence of ResNet and other classic models of deep CNN gradually brought the development of neural networks to the stage of deep networks. The researchers found that as the layers of the network gradually deepened, the representation ability of the model will gradually strengthen, and the possibility of obtaining better generalization ability is greater. However, the deepening of the number of network layers will make the influence of gradient disappearance or gradient explosion more obvious, which makes network training more and more difficult. When the deep neural network is back-propagated, the network parameters are passed to the input layer of the network. In this process, the gradient will be close to zero or tend to infinity, therefore, the deeper the network, the more serious the problem of gradient disappearance and gradient explosion may be, and as the number of layers increases, the CNN may show worse and worse performance on specific datasets, the expressive ability does not continuously improve, which is due to the model degradation caused by the loss of information generated by the deep network.

In 2015, He Kaiming proposed a deep residual network (ResNet) in response to the problem of model degradation in deep networks. ResNet50 is a residual learning model with more network layers than other networks, which can reduce the computational load of the entire network. Figure 3 shows a residual module, which is the basic form of the residual module and is often used in ResNet with a relatively small number of layers. ResNet50 is different from the traditional convolutional neural network. The residual module adds a skip connection between the input and output of the two convolutional layers so that the feature transfer is enhanced. Just like the Skip Connection in Figure 3, it constitutes a layer number fallback mechanism. With the deepening of depth, the training error will not decrease first and then increase. To accelerate the convergence speed, it is necessary

to increase the number of layers of network training to further improve the recognition accuracy.

After the input x is transformed by the convolution layer and the activation layer feature, the output F(x) is obtained, and then by adding the elements of F(x) to the corresponding positions of the elements x, H(x) is obtained in equation (1).

$$\mathrm{H}(\boldsymbol{x}) = \mathrm{F}(\boldsymbol{x}) + \boldsymbol{x} \tag{1}$$

When the result of feature transformation is zero, that is, F(x) = 0 in equation (2).

$$\mathrm{H}(\boldsymbol{x}) = \boldsymbol{x} \tag{2}$$

The network becomes an identity map, the performance of the network will not de-cline when the number of network layers increases, and the model degradation phenomenon is avoided.

C. TRANSFER LEARNING

Transfer learning can be defined in this way, using extant information to conduct different domain problems, and transferring existing knowledge to handle small amounts of samples in the target domain. In the absence of sufficient data to complete the training, the method of transfer learning can enhance the generalization ability of the model itself. The task of learning in the traditional machine learning framework is to learn a classification model on the basis of given sufficient training data. However, huge quantities of training data are hard to get in certain areas. In addition, in practical applications, training dataset and test dataset often do not satisfy the same distribution; transfer learning can solve these problems very well. In practice, this method attempts to transfer as much useful information from the original task as possible to the trained task or scene. The useful information mentioned here can have many manifestations, and it depends on the data. For image data, knowledge can be expressed as generalized features such as texture and color.

In transfer learning based on actual problems, there are mainly four types of connections between the original dataset and the target dataset. If the original dataset and the target dataset have a large degree of similarity, instance-based transfer learning is used. The main purpose is to extract as much as possible useful data for the classification of the target dataset and add it to the target data set. Carry out model training to improve generalization ability. Based on the feature transfer learning in the homogeneous space, the large difference in data is discarded between domains, it can find features at the feature level, and then transfer this feature extracting ability to the target dataset and improve the generalization ability.

In the proposed I-ResNet50, the strategy of weight parameters transferring is applied, and weight parameters trained on ImageNET are transmitted to I-ResNet50. The pre-trained model has strong robustness and generalization ability, freezing network weights and fine-tuning the network to solve the problem of thermal battery defect identification.

III. MODEL ESTABLISHMENT

A. IMPROVED ACTIVATION FUNCTION LReLU

As the most common activation function, the problem of sigmoid local gradient explosion and gradient disappearance are solved by ReLu, which accelerates the convergence of neural networks. The algorithm of ReLu is shown in equation (3).

$$\boldsymbol{y} = \max(0, \boldsymbol{x}) \tag{3}$$

When the input one-dimensional signal is negative, it will discard such input signal, which weakens the model's ability to identify thermal battery defects. Therefore, the method in this paper adopts LReLu as the activation function when the input is negative.

The specific algorithm is as equation (4).

$$y = \begin{cases} x, & x > 0 \\ ax, & x \leq 0 \end{cases} \tag{4}$$

In equation (4), $\boldsymbol{x}$ and $\boldsymbol{y}$ are the input and output of the LReLu activation function, respectively; $\boldsymbol{a}$ is valued according to actual experience. $\boldsymbol{a}$ is set to 0-0.5 by lots of scholars to achieve the best effect. In this study, $\boldsymbol{a}$ is set to 0.01.

B. SENET NETWORK STRUCTURE

The Squeeze and Excite (SENet) network structure is proposed by Hu et al. SENet is to learn the feature weight according to the loss, and the degree of importance of each feature map is obtained, then use this degree of importance to assign a weight value to each feature channel. So the effective feature map has a large weight, the model can achieve better results, and it also does not reduce parameters.

The input feature maps are globally pooled and compressed into a real number by the squeeze operation. The specific algorithm is shown in equation (5).

$$\boldsymbol{F_{sq}}(\boldsymbol{x_i}) = \frac{1}{\boldsymbol{H} \times \boldsymbol{W}} \sum_{\boldsymbol{a}}^{\boldsymbol{H}} \sum_{\boldsymbol{b}}^{\boldsymbol{W}} \boldsymbol{x_i}(\boldsymbol{a}, \boldsymbol{b}) \tag{5}$$

In equation (5), $\boldsymbol{x_i}$ represents the i-th feature map of input size $\boldsymbol{H} \times \boldsymbol{W}$.

The excitation operation is mainly composed of two FC and two activation functions, which can help capture the channel correlation and generate the weight of the corresponding channel. The algorithm is shown in equation (6).

$$\boldsymbol{y_i} = \boldsymbol{F_{ex}}(\boldsymbol{F_{sq}}(\boldsymbol{x_i}), \omega) = \sigma(\omega_2 \delta(\omega_1 \boldsymbol{F_{sq}}(\boldsymbol{x_i}))) \tag{6}$$

In Equation (6), ω_1 represents the first FC calculation; ω_2 represents the second FC calculation, the focus parameter; $\boldsymbol{F_{sq}}(\boldsymbol{x_i})$ represents the output value after the Squeeze operation; δ represents the activation function ReLU; σ represents the sigmoid function; the specific algorithm is as equation (7).

$$\sigma = \frac{1}{1 + \boldsymbol{e}^{-x}} \tag{7}$$

In Equation (7), $\boldsymbol{x}$ represents the output value after two full connection calculations.

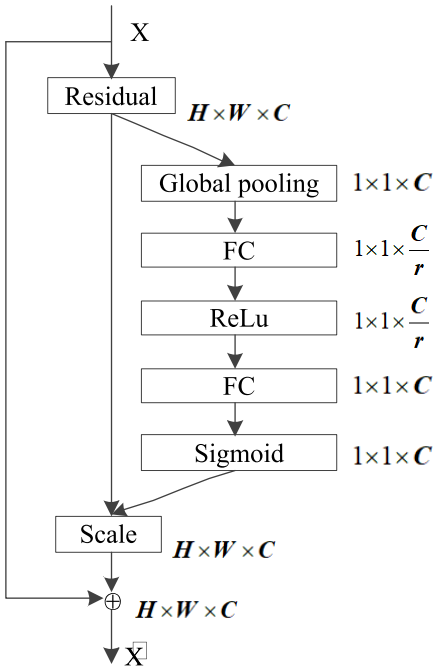


FIGURE 4. Schematic structure of embedding SENet into the residual network.

The scale operation is to multiply the weight coefficients of each channel learned earlier with all elements of the corresponding channel by multiplication, so as to enhance the important features and be more directional. Figure 4 is a schematic structure of embedding SENet into the residual network.

C. MODEL STRUCTURE

Transfer learning and SENet network are introduced to establish a thermal battery defect identification model I-ResNet50 based on Res-Net50 network.

The weight parameters pre-trained on ImageNet are transferred to the I-ResNet50 network model. The parameters of the bottleneck layer are initialized by the pre-training model at the beginning of experiments; the different Block features of ResNet50 are transferred to the I-ResNet50 recognition model. Then freeze the other blocks in turn for training tests, to better verify the classification performance of feature transfer schemes, five categories of feature transfer schemes are set up.

In ImageNet, each synset provides an average of 1000 images. Each concept image is quality-controlled and human-annotated. At present, there are a total of 14,197,122 images in ImageNet, which are divided into 21,841 categories (synsets). The major categories include fabric flower, fruit, fungus, furniture, geometric formation, musical instrument, plant, reptile, structure, person, etc. It is useful for the I-ResNet50 network learning defect image characteristic of the thermal battery. ImageNet dataset is shown in Figure 5.

In the resnet50 model, a block implementation consists of 1×1, 64, 3×3, 64, 1×1, 256, etc. That is when defining

FIGURE 5. ImageNet dataset.

Bottleneck (planes, stride, down sample), the change may be the size of the convolution kernel. For each layer, channels become 4 planes after each block, so the number of input channels for the next block should be 4 planes. The first layer has a total of 3 blocks, each block has 3 layers, a total of 9 layers, all of which are not affected by stride, that is, stride=1, the feature map is not changed; The second layer has a total of 4 blocks, each block has 3 layers, a total of 12 layers. The second layer of the first block is affected by stride, which may change the feature map, the feature map is not affected by the remaining 11 layers; the third layer has a total of 6 blocks, and each block has 3 layers, a total of 18 layers. The second layer of the first block is affected by stride, which may change the feature map, the feature map is not changed by the remaining 17 layers. The fourth layer is similar to the above.

Figure 6(b) shows the architecture diagram of the I-ResNet50 network.

The thermal battery assembly defect images of the dataset are used as the input of the I-ResNet50 network. First, a 7×7 convolution operation is performed. After that, batch normalization is performed, which is activated by the activation function LReLu, after a pooling operation with 3×3, 4 identities residual block learning is performed, among them, the order of convolution kernels in each residual block of I-ResNet50 is 1×1, 3×3 and 1×1. 1×1 can reduce the number of channels of the input feature, and the last layer of 1×1 can increase channels of output features, in this way, the number of channels before and after passing through the bottleneck layer is guaranteed to be consistent, and the calculation amount of the 3×3 convolution kernel in the middle layer is reduced.

After I-ResNet50 completes the convolution calculation, four designed residual blocks are used, and each identity residual block in I-ResNet50 is stacked 3, 4, 6, and 3 times in a turn. Then the average pooling operation is performed, and the multi-dimensional vector output is converted into a one-dimensional vector output. FC1 is 1000, FC2 is 4, dropout rate is 0.5. Half of the parameters generated by the

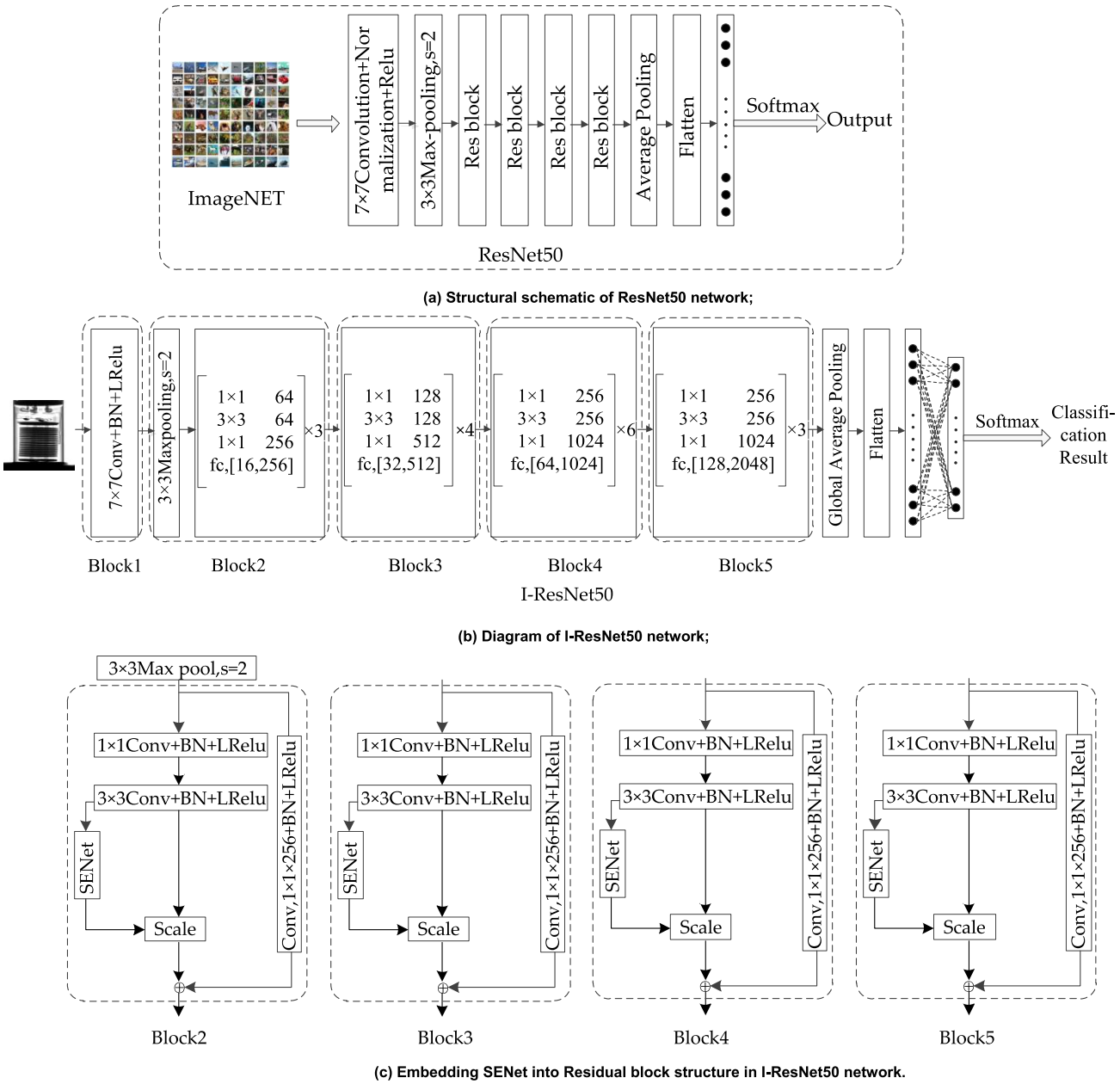


FIGURE 6. I-ResNet50 network transfer learning architecture diagram.

fully connected layer are randomly discarded, and then the LReLu function is to alleviate gradient dispersion. Lastly, the prediction result of the thermal battery assembly defect image is output through the Softmax function.

In Figure 6(c) SENet is embedded into Residual block structure in the I-ResNet50 network. The dimension information represents output. GAP is used as the Squeeze operation and outputs the same weights. It has two advantages: (1) It can fit correlation between channels; (2) The computation is greatly reduced.

Then, normalized weights between 0 and 1 are obtained through the sigmoid function, and a Scale operation is used to weight the normalized weight. Figure 6 is the final network architecture diagram.

D. ADAM OPTIMIZER

The Adam optimizer is proposed by two scholars in 2014, and it performs first-order gradient optimization on a random objective function, combining the advantages of two optimization algorithms AdaGrad and RMSProp, it calculates step size and dynamically adjusts the learning rate of each parameter by the first-order and second-order moment estimation. Each iterative learning has a clear range, which makes the parameters change smoothly, and also ensures that the possibility of gradient explosion or gradient disappearance is suppressed. The following equations are derived as the parameter update principle of the Adam optimizer.

$$\boldsymbol{m}_{\boldsymbol{w}}^{t+1} = \beta_1 \boldsymbol{m}_{w}^{t} + (1 - \beta_1)\nabla \boldsymbol{L}^{t} \tag{8}$$

In Equation (8): $\boldsymbol{m}$ represents the first-order moment estimation, β represents the hyper parameter, and $\boldsymbol{t}$ represents the number of updates.

$$\boldsymbol{v}_{\boldsymbol{w}}^{\boldsymbol{t}+1} = \beta_2 \boldsymbol{m}_{\boldsymbol{w}}^{\boldsymbol{t}} + (1-\beta_2)(\nabla \boldsymbol{L}^{\boldsymbol{t}})^2 \tag{9}$$

$\boldsymbol{v}$ represents the second-order moment estimation.

$$\hat{\boldsymbol{m}}_{\boldsymbol{w}} = \frac{\boldsymbol{m}_{\boldsymbol{w}}^{\boldsymbol{t}+1}}{1-\beta_2^{t+1}} \tag{10}$$

$$\hat{\boldsymbol{v}}_{\boldsymbol{w}} = \frac{\boldsymbol{v}_{\boldsymbol{w}}^{\boldsymbol{t}+1}}{1-\beta_2^{t+1}} \tag{11}$$

Equation (10) and Equation (11) are combined to obtain Equation (12), which $\boldsymbol{w}$ represents the parameter to be updated.

$$\boldsymbol{w}^{\boldsymbol{t}+1} \leftarrow= \boldsymbol{w}^{\boldsymbol{t}} - \eta \frac{\hat{\boldsymbol{m}}_{\boldsymbol{w}}}{\sqrt{\hat{\boldsymbol{v}}_{\boldsymbol{w}}} + \varepsilon} \tag{12}$$

E. SGD OPTIMIZER

In the SGD algorithm, the iterative update formulas of parameters are equation (13) and equation (14).

$$\mathrm{W} = \mathrm{W} - \alpha \mathrm{dW} \tag{13}$$

$$\boldsymbol{b} = \boldsymbol{b} - \alpha \mathrm{d}\boldsymbol{b} \tag{14}$$

After the Momentum is introduced, the iterative update formula is equation (15).

$$\begin{aligned} \boldsymbol{v}_{\mathrm{dW}} &= \beta \boldsymbol{v}_{\mathrm{dW}} + (1-\beta)\mathrm{dW} \\ \boldsymbol{v}_{\mathrm{db}} &= \beta \boldsymbol{v}_{\mathrm{db}} + (1-\beta)\mathrm{db} \\ \mathrm{W} &= \mathrm{W} - \alpha \boldsymbol{v}_{\mathrm{dw}}, \quad \boldsymbol{b} = \boldsymbol{b} - \alpha \boldsymbol{v} \end{aligned} \tag{15}$$

In equation (15), W represents the weight, $\boldsymbol{b}$ represents the bias value, α (learning rate) and β are hyper parameters, and the value β determines the influence of the last update value on this update.

For sparse data or features, sometimes it may be necessary to update faster for features that appeared infrequently. At this time, SGD cannot meet the requirements. But Momentum can also be introduced, the last updated value is taken into account on each update of the network parameters, and strengthen the update of the same parameters with the previous gradient direction, which can increase stability and improve learning speed so that the network can get rid of the local optimal solution to a certain extent.

F. BATCH NORMALIZATION

Batch normalization is also one of the main optimization methods of neural net-works. Batch normalization refers to normalizing w*x+b according to the features in front of the activation function in the neural network. There are three advantages as follows:

(1) It can improve the flow of gradients in the network. Normalization can scale all features to [0, 1] so that the gradient during back propagation is around 1, and the phenomenon of gradient disappearance is avoided.

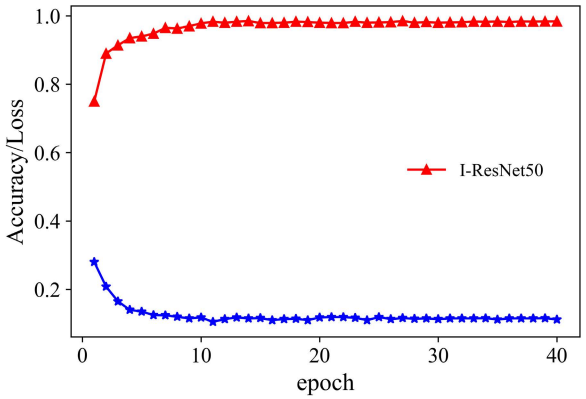


FIGURE 7. Accuracy and loss of when transferring bottleneck layer features.

(2) It can increase the learning rate. The normalized data can reach convergence quickly.

(3) It can reduce the dependence of model training on initialization.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. EXPERIMENTAL COMPARISION OF DIFFERENT TRANSFERRING SCHEMES

First, experiment 1 is carried out, the bottleneck layer features of the ResNet50 net-work model are transferred to the I-ResNet50 network model proposed in this paper, and defect image recognition is performed under the enhanced thermal battery image data set. Batch size is 32, learning rate decay value is 0.9, and iterations are 40.

L2 regularization is added to Adam in the experiment to prevent overfitting and improve generalization ability. L2 parameter regularization is also known as weight de-cay, as in Equation (16);

$$\Omega(\omega) = \|\omega\|^2 \tag{16}$$

The regular term makes the weight decay closer to the origin, and the network model becomes simpler, so as to achieve the purpose of optimizing the parameters. This experiment takes L2=0.0001 for the experiment. The experimental results of the I-ResNet50 network model are obtained, as shown in Figure 7.

Secondly, experiment 2 is designed to verify the feature transfer scheme in this paper by the I-ResNet50 network model. According to the I-ResNet50 network model structure, the different Block features of ResNet50 are transferred to the I-ResNet50 recognition model. To better verify the classification performance of feature transfer schemes, five categories of feature transfer schemes are set up. Table 1 lists the five categories of feature transfer schemes for the comparative experiments.

Among them, Fixed represents freezing the Block, the corresponding weight pa-rameters of the trained model are transferred to the frozen Block, and Train represents retraining the Block. FC1 is 1000; FC2 is the number of sample categories

TABLE 1. Five categories of feature transfer schemes.

Plan	Bolck1	Bolck2	Bolck3	Bolck4	Bolck5	Top
Plan_5	Fixed	Fixed	Fixed	Fixed	Fixed	Train
Plan_4	Fixed	Fixed	Fixed	Fixed	Train	Train
Plan_3	Fixed	Fixed	Fixed	Train	Train	Train
Plan_2	Fixed	Fixed	Train	Train	Train	Train
Plan_1	Fixed	Train	Train	Train	Train	Train

TABLE 2. The experimental results of the five categories of feature transfer schemes.

Plan	Training		Validation	
	Loss	Accuracy	Loss	Accuracy
Plan_5	0.236	91.5	0.313	90.9
Plan_4	0.125	98.1	0.164	95.8
Plan_3	0.130	97.7	0.181	95.3
Plan_2	0.162	95.9	0.249	92.9
Plan_1	0.171	95.1	0.278	91.7

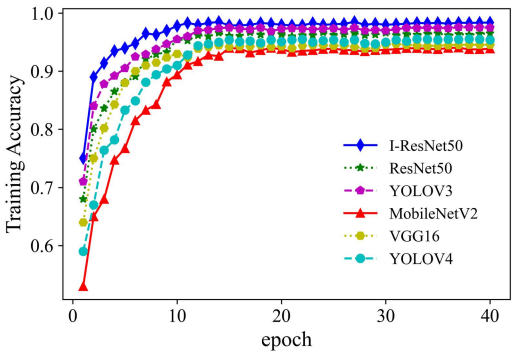
4. The activation functions are both LReLu functions. The model optimizer continues to use the Adam optimizer. The experimental results of the five categories of feature transfer schemes are shown in Table 2.

Experimental classification results obtained by using the original training model to perform feature transfer on image samples of a small data set are more accurate than re-starting training a new neural network by 10 percentage points. At the same time, it can be seen that it's not the more retrained layers, the fewer the frozen layers, and the better the recognition effect. Plan_4 and plan_3 achieve better results in these five schemes. Retrained Block5 and the top layer of plan_4 have higher recognition accuracy than re-trained Block4, Block5, and the top layer of plan_3.

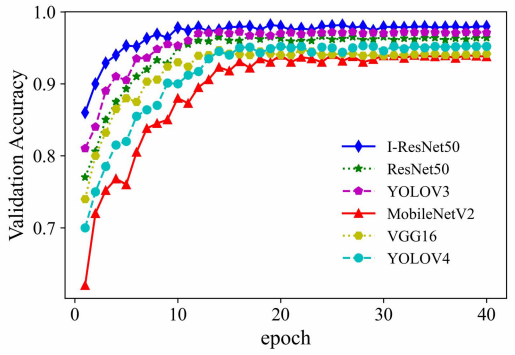
B. ANALYSIS OF RECOGNITION RATE

When the data is almost balanced, the accuracy rate can fully evaluate the performance of the model. Experiment 3 is a comparison experiment of accuracy and loss. The results of the I-ResNet50 model and the other five network models on the training set and validation set of thermal battery assembly defect images are compared. The other five network models are ResNet50, YOLOV3, MobileNetV2, VGG16, and YOLOV4, respectively.

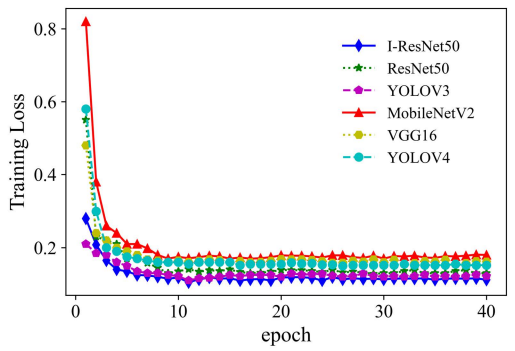
Figure 8 is a graph of the accuracy and loss results of the six networks. Each of the six networks ran iteratively for 40 epochs. From the accuracy of the training set, it can be found that the I-ResNet50 network has the best recognition effect on thermal battery defect images, its accuracy can reach 98.1%, and the MobileNetV2 network has the worst recognition effect. The accuracy of I-ResNet50 is nearly 5% higher than



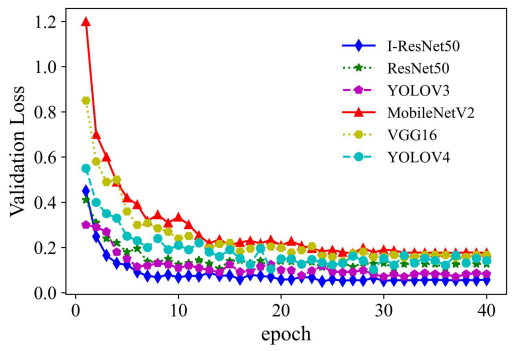
(a) Training accuracy;



(b) Validation accuracy;



(c) Training loss;



(d) Validation loss;

FIGURE 8. Accuracy and loss during epochs.

the MobileNetV2 network. The ResNet50 network reached 96.26%, the YOLOV3 network reached 97.15%, the VGG16 network reached 94.31%, the MobileNetV2 network reached

TABLE 3. Accuracy, recall, and F1-score of six models.

Model	Precision(%)				Recall(%)				F1-score(%)				T(min)
	Q-A	M-C-P	W-N-S	R-S	Q-A	M-C-P	W-N-S	R-S	Q-A	M-C-P	W-N-S	R-S	
I-ResNet50	99.6	98.1	99.1	95.4	99.6	95.1	99.5	98.1	99.6	96.6	99.3	97.5	28
ResNet50	99.3	94.6	98.0	92.9	98.9	92.2	99.1	94.9	99.1	93.4	98.5	93.9	62
YOLOV3	99.5	96.0	98.9	94.2	99.3	94.3	98.7	96.3	99.4	95.2	98.8	95.2	57
MobileNetV2	97.3	92.9	96.7	87.4	98.0	86.7	96.7	92.8	97.6	89.8	96.7	90.1	35
VGG16	97.1	93.8	96.7	89.5	98.6	88.4	96.9	93.2	97.8	91.1	96.8	91.4	92
YOLOV4	98.4	95.2	98.4	90.0	99.4	89.2	98.4	94.9	98.9	96.8	98.4	92.5	49

TABLE 4. Experimental results of different optimizers and learning rates.

Learning rate	Optimization		
	SGD	Adam	RMSprop
0.01	97.6	96.3	97.3
0.001	97.9	98.1	96.1
0.0001	88.5	97.9	97.8

93.56%, and the YOLOV4 network reached 95.51%. After 40 epochs of training, all six networks have fully converged. The I-ResNet50 network has been close to complete convergence from the 12th epoch with the fastest convergence and the smallest loss, Loss value reached 0.13. However, the loss value curve of the MobileNetV2 network and the VGG16 network on the validation set fluctuates greatly.

This shows that the effective transmission of feature information is strengthened by improving the convolution, BN, and SENet module in the network structure, the feature expression ability of the bottleneck structure is enhanced, and the difficulty of network training is reduced. By improving the residual structure, the problem of losing a large amount of important information of the input feature map in the sampling operation of the residual branch is effectively alleviated, so that the network can make full use of the learned features. Compared with the other five networks, the I-ResNet50 network has a higher average recognition rate and faster convergence.

Precision, Recall, F1-score, T(Network Running Time) are used as evaluation indicators of I-ResNet50 to reflect the improvement effect of I-ResNet50 and evaluate the effectiveness of I-ResNet50. First, make the following settings:

True Positive (TP): The I-ResNet50 model correctly predicts the true thermal battery defect categories;

True Negative (TN): The I-ResNet50 model predicts other thermal battery defect categories as non-true defect categories;

False Positive (FP): The I-ResNet50 model incorrectly predicts other thermal battery defect categories as true defect categories;

False Negative (FN): The I-ResNet50 model mispredicts the true thermal battery defect categories as other defect categories.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{17}$$

$$Precision = \frac{TP}{TP + FP} \tag{18}$$

$$Recall = \frac{TP}{TP + FN} \tag{19}$$

$$F1 - score = 2\frac{Precison \cdot Recall}{Precison + Recall} \tag{20}$$

The experimental data of the six networks are shown in Table 3. Four categories of thermal battery images include Qualified Assembly (Q-A), Missing Current Plate (M-C-P), Wrong Number of Stacks (W-N-S), and Reverse Stack (R-S) images.

The I-ResNet50 network's precision of the four categories of thermal battery defect images reached 99.64%, 98.17%, 99.11%, and 95.40% respectively; its accuracy reached 98.10%, which is the most effective model compared to ResNet50, YOLOV3, MobileNetV2, VGG16, and YOLOV4. It can be found that from the experimental data, the running time of I-ResNet50 is 28 minutes, which is the shortest. Although the running time of I-ResNet50 is similar to MobileNetV2, the accuracy of MobileNetV2 network is too low, so the I-ResNet50 network proposed has the best effect, which further illustrated the effectiveness of the CNN deep learning algorithm and knowledge transfer strategy.

C. EXPERIMENTAL COMPARISION OF DIFFERENT OPTIMIZERS AND LEARNING RATES

There are many optimization methods that can optimize various neural network models. The ultimate purpose of each model is to find the optimal solution. The selection of different optimizers and parameter settings will also cause different training effects. In the training process, if the learning rate is too large, the convergence curve will oscillate, so it is not easy to obtain the optimal value.

To verify the effect of different optimizers and learning rates on the hot battery defect identification results, the optimal experimental scheme in Experiment 2 is used, Block5

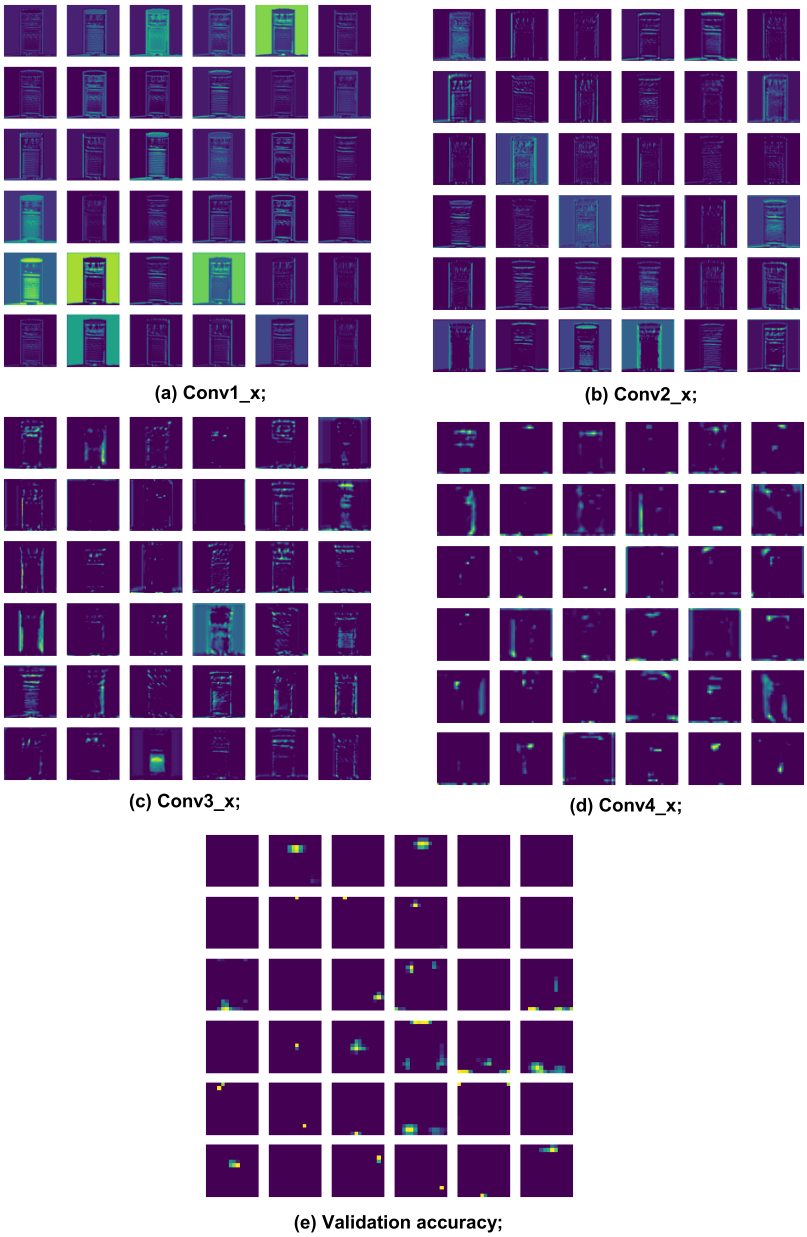


FIGURE 9. Convolution kernel visualization of thermal battery image.

and the top layer of plan_4 is retrained, The SGD, Adam, and RMSprop optimizer are used to test on the improved model I-ResNet50 network model. The parameters of the three optimizers are set: (1) SGD: Batch size is 32, the momentum is 0.9, iterations are 45. (2) Adam: Batch size is 32, learning rate decay value is 0.9, and iterations are 45. (3) RMSprop: rho is 0.9, epsilon is 1e-06, iterations are 45.

The result of the last iteration of all experiments is taken as the final result. The initial learning rates are set as 0.01, 0.001, and 0.0001 respectively. Test results are shown in Table 4.

The experimental results show that when the optimizer is Adam and the learning rate is 0.001, the model can converge rapidly, and the test accuracy rate can reach 98.1%, which is more suitable for the training of thermal battery defect identification.

D. MODEL FEATURE EXTRACTION AND VISUALIZATION

The convolution kernel of the I-ResNet50 can extract features of different layers of thermal battery defect images and output the feature map of this layer. By visualizing the feature map of the I-ResNet50 network model, it can be visually observed how the I-ResNet50 network model transforms the hot battery defect image during the training process. And the actual meaning of the convolution kernel can be further understood so that the working principle of the I-ResNet50 network model can be better understood.

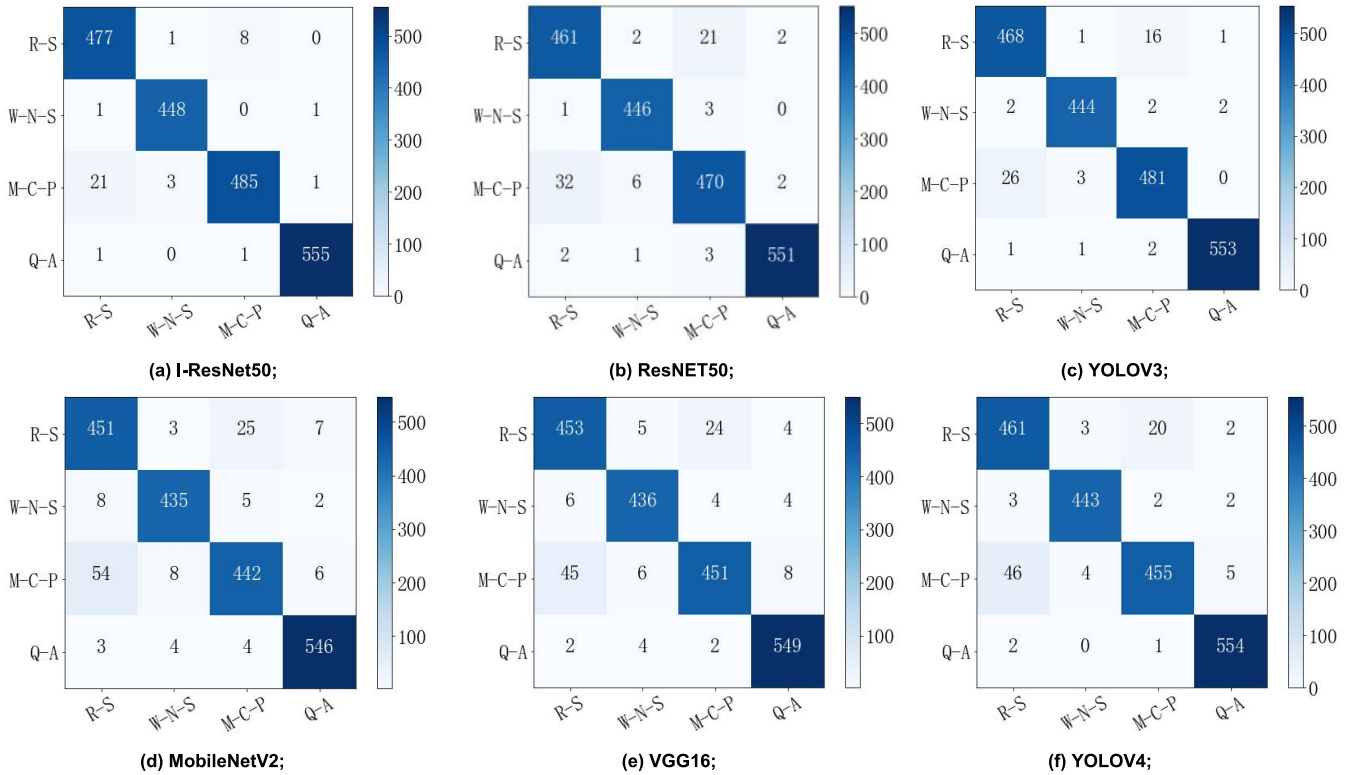


FIGURE 10. Confusion matrix for six models to identify thermal battery defect images.

According to the I-ResNet50 network model structure, I-ResNet50 consists of 5 groups of 50 convolutional layers, and each convolutional layer performs feature extraction on the thermal battery defect image. In order to see the model extraction effect more intuitively, taking a thermal battery defect image as an example, the I-ResNet50 model is used for training, and the output of the convolution kernel of each layer of the model is visualized. Figure 9 is the result after visualization of the thermal battery defect image feature map.

Figure 9(a) and Figure 9(b) are the output feature maps of the Conv1_x and Conv2_x layers of the thermal battery defect image. From Figure 9(a) and Figure 9(b), it can be concluded that the contour features of the thermal battery defect image is displayed through different color channels, which learned by convolution kernel is contour edge characters of thermal battery, all the characters of the image is retained by Conv1_x and Conv2_x layers.

Figure 9(c) and Figure 9(d) are the output feature maps of Conv3_x and Conv4_x layers. In the feature map, the outline is more obvious, and the resolution is getting smaller. As the number of layers deepens, the content of the output of the convolution kernel becomes more and more abstract and retained features become less and less.

Figure 9(e) is shown that the output feature map of the Conv5_x layer becomes blurred and more blank content appears. Some regions are found to be easier to distinguish than others due to the extraction of the strongest features.

It can be found the shape and surface features of the thermal battery defect image are extracted, and features are more representative so that the strongest features of thermal battery defects are extracted. It can be concluded that CNN can form new high-level abstract features through the integration of low-level features, and use high-level abstract features to distinguish the essential features of objects.

E. ANALYSIS OF CONFUSION MATRIX

After testing, the confusion matrix results of six networks for identifying thermal battery defect images are shown in Figure 10. Among them, 2003 thermal battery defect image samples are tested by the improved I-ResNet50 network, and the number of correct samples is 1965, and the overall accuracy is 98.10%.

As can be seen from Figure 10(a), among the 486 Reverse Stack samples, 1 sample is misjudged as Wrong Number of Stacks, and 8 samples are misjudged as Missing Current Plate; among the 450 Wrong Number of Stacks samples, 1 sample is misjudged as Reverse Stack, and 1 sample is misjudged as Qualified Assembly; among the 510 Missing Current Plate samples, 21 samples are misjudged as Reverse Stack, 3 samples are misjudged as Wrong Number of Stacks, and 1 sample is misjudged as Qualified Assembly; among the 557 Qualified Assembly samples, 1 sample is misjudged as Reverse Stack, and 1 sample is misjudged as Missing Current Plate.

Compared to the confusion matrices, it is concluded the number of Qualified Assembly samples correctly identified as Qualified Assembly is 555, 551, 553, 546, 549, and 554. The accuracy of the I-ResNet50 model reached 99.64%, which is the highest. The number of 486 Reverse Stack samples correctly identified as Reverse Stack is 477, 461, 468, 451, 453, 461, the precision of the I-ResNet50 model reached 95.4%, and the precision of the MobileNetV2 model reached 87. 4%, this is the lowest precision among the six models.

The reasons for the inaccurate identification are as follows:

1 The number of defect images in the thermal battery dataset is not large enough; the defect features learned by the network model are limited.
2 It is possible that the Reverse Stack is identified as Qualified Assembly because the defect parts in the two defect images are the same after the rotation operation, which leads to the recognition error of the model.
3 The texture similarity of Reverse Stack and Missing Current Plate makes it difficult to correctly identify by the model.

It can be seen that although the errors of Reverse Stack and Missing Current Plate occur between pictures with similar texture features, the errors are within an acceptable range; this further verifies the reliability and high recognition rate of I-ResNet50. Com-pared with the five models of ResNet50, YOLOV3, MobileNetV2, VGG16, and YOLOV4, the improved I-ResNet50 network model confusion matrix shows the best performance, which provides an important method for correctly identifying defect image features in the thermal battery dataset.

V. CONCLUSION

Aiming at the problems of insufficient capability, time-consuming and labor-intensive manual diagnosis, and low accuracy of traditional diagnostic methods for thermal battery defects, an improved residual neural network thermal battery defect identification method based on transfer learning is proposed. The conclusions drawn from this paper are as follows:

(1) This paper constructs an improved deep residual neural network thermal battery defect diagnosis model I-ResNet50 based on a residual neural network combined with SENet. By adding the SENet network structure to the residual block, the model focuses on the feature channels related to the target and weakens other irrelevant feature channels to improve the performance of the model. By using global average pooling instead of FC layers, the training speed is accelerated. In the training process, in order to suppress overfitting, data enhancement processing is performed on the training samples. At the same time, fine-tuning operations such as the LReLU activation function and Dropout are used to enhance recognition ability.

(2) The I-ResNet50 network model is experimentally verified on the enhanced defect image dataset, and the overall accuracy of defect recognition reaches 98.10%. Compared with ResNet50, YOLOV3, MobileNetV2, VGG16, and YOLOV4 five network models, under the same conditions, the I-ResNet50 network model converges faster and the computational cost is greatly reduced. The effectiveness of the I-ResNet50 network model is proved through four experiments, and it has good anti-interference ability and generalization ability.

In the future work, it is planned to further improve the model structure based on the residual network structure, and more optimization methods and parameter adjustment methods are used to establish a new model to identify various defect images of multiple categories of thermal batteries.

ACKNOWLEDGMENT

The editor and anonymous reviewers for providing useful suggestions for improving this article.

AUTHOR CONTRIBUTIONS

Methodology, Wenchao Xu; software, Wenchao Xu; writing—draft preparation, Wenchao Xu; writing—review and editing, Wenchao Xu; and supervision, Sixiang Zhang, Fang Bai, and Tao Zhao.

DATA AVAILABILITY STATEMENT

All data included in this study are available upon request by contact with the corresponding author.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] S. Narayanan, X. Li, S. Yang, H. Kim, A. Umans, I. S. McKay, and E. N. Wang, "Thermal battery for portable climate control," *Appl. Energy*, vol. 149, pp. 104–116, Jul. 2015.
[2] Y. Niu, Z. Wu, J. Du, and W. Duan, "A new thermal battery for powering borehole equipment: The discharge performance of Li–Mg–B alloy/$LiNO_3$–KNO_3/MnO_2 cells at high temperatures," *J. Power Sources*, vol. 245, pp. 537–542, Jan. 2014.
[3] M. Khalil Anwar, B. S. Yilbas, and S. Z. Shuja, "A thermal battery mimicking a concentrated volumetric solar receiver," *Appl. Energy*, vol. 175, pp. 16–30, Aug. 2016.
[4] M. H. Miles, "Exploration of molten hydroxide electrochemistry for thermal battery applications," *J. Appl. Electrochemistry*, vol. 33, no. 11, pp. 1011–1016, 2003.
[5] D. Reisner, T. D. Xiao, H. Ye, J. Dai, Ronald A. Guidotti, and F. W. Reinhardt, "Thermal-sprayed thin film cathodes for thermal battery," *J. New Mater. Electrochemical Syst.*, vol. 2, no. 4, pp. 279–283, 1999.
[6] S. Kumar and P. Pandey, "FPGA implementation of image segmentation by using edge detection based on Sobel edge operator," *Int. J. Res. Eng. Technol.*, vol. 2, no. 10, pp. 198–203, Oct. 2013.
[7] Y.-I. Ohta, T. Kanade, and T. Sakai, "Color information for region segmentation," *Comput. Graph. Image Process.*, vol. 13, no. 3, pp. 222–241, 1980.
[8] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. SMCS-9, no. 1, pp. 62–66, Feb. 1979.
[9] X.-C. Yuan, L.-S. Wu, and Q. Peng, "An improved Otsu method using the weighted object variance for defect detection," *Appl. Surf. Sci.*, vol. 349, pp. 472–484, Sep. 2015.
[10] S. A. Hojjatoleslami and J. Kittler, "Region growing: A new approach," *IEEE Trans. Image Process.*, vol. 7, no. 7, pp. 1079–1084, Jul. 1998.
[11] P. C. Chen and T. Pavlidis, "Segmentation by texture using a co-occurrence matrix and a split-and-merge algorithm," *Comput. Graph. Image Process.*, vol. 10, no. 2, pp. 172–182, Jun. 1979.

[12] Y. Huang, F. Zhang, M. Yu, and Y. Zhao, "Silicon wafer defect extraction based on morphological filter and watershed algorithm," in *Proc. Int. Conf. Comput. Sci. Softw. Eng.*, 2008, pp. 141–144.
[13] T.-C. Su, M.-D. Yang, T.-C. Wu, and J.-Y. Lin, "Morphological segmentation based on edge detection for sewer pipe defects on CCTV images," *Expert Syst. Appl.*, vol. 38, no. 10, pp. 13094–13114, Sep. 2011.
[14] Z. Zheng, H. Wang, and E. K. Teoh, "Analysis of gray level corner detection," *Pattern Recognit. Lett.*, vol. 20, no. 2, pp. 149–162, Feb. 1999.
[15] L. Yi-Bo and L. Jun-Jun, "Harris corner detection algorithm based on improved contourlet transform," *Proc. Eng.*, vol. 15, pp. 2239–2243, Jun. 2011.
[16] Y. Chen, M. Xu, H.-L. Liu, W.-N. Huang, and J. Xing, "An improved image mosaic based on Canny edge and an 18-dimensional descriptor," *Optik*, vol. 125, no. 17, pp. 4745–4750, Sep. 2014.
[17] J. Ning, L. Zhang, D. Zhang, and C. Wu, "Robust object tracking using joint color-texture histogram," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 23, pp. 1245–1263, Nov. 2009.
[18] H. R. Li, T. Jiang, and K. Zhang, "Efficient and robust feature extraction by maximum margin criterion," *IEEE Trans. Neural Netw.*, vol. 17, no. 1, pp. 157–165, Feb. 2006.
[19] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.
[20] Z. Liu, B. Wang, C. Li, M. Yu, and S. Ding, "Fabric defect detection based on deep-feature and low-rank decomposition," *J. Eng. Fibers Fabrics*, vol. 15, Jan. 2020, Art. no. 155892502090302.
[21] M. Sharma and G. Kaur, "Integrated approach for defect detection in ceramic tiles," *Int. J. Comput. Technol.*, vol. 3, no. 2, pp. 463–468, 2012.
[22] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Math. Comput.*, vol. 19, no. 90, pp. 297–301, 1965.
[23] D.-M. Tsai and T.-Y. Huang, "Automated surface inspection for statistical textures," *Image Vis. Comput.*, vol. 21, no. 4, pp. 307–323, 2003.
[24] W.-C. Li and D.-M. Tsai, "Wavelet-based defect detection in solar wafer images with inhomogeneous texture," *Pattern Recognit.*, vol. 45, no. 2, pp. 742–756, 2012.
[25] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[26] L. Fu, Y. Feng, Y. Majeed, X. Zhang, J. Zhang, M. Karkee, and Q. Zhang, "Kiwifruit detection in field images using Faster R-CNN with ZFNet," *IFAC-PapersOnLine*, vol. 51, no. 17, pp. 45–50, 2018.
[27] N. Shibata, M. Tanito, K. Mitsuhashi, Y. Fujino, M. Matsuura, H. Murata, and R. Asaoka, "Development of a deep residual learning algorithm to screen for glaucoma from fundus photography," *Sci. Rep.*, vol. 8, no. 1, pp. 1–9, Dec. 2018.
[28] N. Neogi, D. K. Mohanta, and P. K. Dutta, "Review of vision-based steel surface inspection systems," *EURASIP J. Image Video Process.*, vol. 2014, no. 1, pp. 1–19, 2014.
[29] S. Wang, X. Xia, L. Ye, and B. Yang, "Automatic detection and classification of steel surface defect using deep convolutional neural networks," *Metals*, vol. 11, no. 3, p. 388, Feb. 2021.
[30] W. Chen, Y. Gao, L. Gao, and X. Li, "A new ensemble approach based on deep convolutional neural networks for steel surface defect classification," *Proc. CIRP*, vol. 72, pp. 1069–1072, Jan. 2018.
[31] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.
[32] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
[33] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, May 2016, pp. 779–788.
[34] W. Ma, X. Wang, and J. Yu, "A lightweight feature fusion single shot multibox detector for garbage detection," *IEEE Access*, vol. 8, pp. 188577–188586, 2020.
[35] Y. Zhang, W. Zhou, Y. Wang, and L. Xu, "A real-time recognition method of static gesture based on DSSD," *Multimedia Tools Appl.*, vol. 79, nos. 25–26, pp. 17445–17461, Jul. 2020.
[36] X. Liang, J. Zhang, L. Zhuo, Y. Li, and Q. Tian, "Small object detection in unmanned aerial vehicle images using feature fusion and scaling-based single shot detector with spatial context analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 6, pp. 1758–1770, Jun. 2020.
[37] 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 Trans. Instrum. Meas.*, vol. 67, no. 2, pp. 257–269, Feb. 2017.
[38] G. Kang, S. Gao, L. Yu, and D. Zhang, "Deep architecture for high-speed railway insulator surface defect detection: Denoising autoencoder with multitask learning," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 8, pp. 2679–2690, Aug. 2019.
[39] W.-H. Wu, J.-C. Lee, and Y.-M. Wang, "A study of defect detection techniques for metallographic images," *Sensors*, vol. 20, no. 19, p. 5593, Sep. 2020.
[40] T. Zhao, "Research on defects detecting method in thermal battery based on a comparison with coordinates of the peaks in gray histogram," *J. Electron. Meas. Instrum.*, vol. 34, pp. 133–140, Sep. 2020.

WENCHAO XU (Member, IEEE) received the M.I.S. degree from the Hebei University of Technology, China, in 2009. He is currently a Lecturer at the College of Information Engineering, Tianjin University of Commerce, China. His research interests include pattern recognition, intelligent control, and deep learning.

SIXIANG ZHANG (Member, IEEE) received the B.Sc. degree from Tianjin University, in 1982, the M.Sc. degree from Zhejiang University, in 1993, and the Ph.D. degree from Tianjin University, in 1996. He is currently a Professor and a Ph.D. Supervisor at the Hebei University of Technology. His research interests include the theory and technology of mechanical geometric measurement and computer image processing technology.

FANG BAI received the M.Sc. degree from the Hebei University of Technology, in 2009. She is currently a Lecturer at the Tianjin University of Commerce, China. Her research interests include measurement technology and instruments and deep learning.

TAO ZHAO (Member, IEEE) received the M.Sc. degree from the University of Science and Technology Beijing, in 2012. He is currently pursuing the Ph.D. degree with the Hebei University of Technology. He is also an Associate Professor at the Tianjin University of Technology, China. His research interests include measurement technology and instruments and deep learning.