

## *Retraction*

# **Retracted: Image Analysis Method of Substation Equipment Based on Convolutional Neural Network**

### **Security and Communication Networks**

Received 26 December 2023; Accepted 26 December 2023; Published 29 December 2023

Copyright © 2023 Security and Communication Networks. 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.

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.


The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

### **References**

- [1] J. Zhang, Y. Liu, L. Yuan, and H. Jia, "Image Analysis Method of Substation Equipment Based on Convolutional Neural Network," *Security and Communication Networks*, vol. 2022, Article ID 1718514, 10 pages, 2022.

## Research Article

# Image Analysis Method of Substation Equipment Based on Convolutional Neural Network

Jingjing Zhang <sup>1</sup>, Yuxin Liu,<sup>1</sup> Lin Yuan,<sup>2</sup> and Haowei Jia<sup>3</sup>

<sup>1</sup>School of Electrical Engineering, Zhengzhou Railway Vocational & Technical College, Zhengzhou 450052, China

<sup>2</sup>Beijing Institute of Science and Technology, China Railway Beijing Group Co., Ltd, Beijing 100081, China

<sup>3</sup>Power Supply Maintenance Technology Center,

Zhengzhou East High Speed Rail Infrastructure Section of China Railway Zhengzhou Group Co., Ltd, Zhengzhou 450052, China

Correspondence should be addressed to Jingjing Zhang; zhangjingjing@zzrvtc.edu.cn

Received 20 June 2022; Revised 18 July 2022; Accepted 29 July 2022; Published 27 August 2022

Academic Editor: Hangjun Che

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

Due to the continuous development of computer technology to promote the continuous progress of substation automation technology, the current substation equipment is diverse and there are many interferences, making the accuracy of the image processing algorithm to be low, and there is a lack of a complete automatic processing system. Convolutional neural networks (CNNs) are one of the most important breakthroughs in artificial intelligence in the last decade, especially in the field of image recognition, and have made important research achievements. In this study, we apply CNNs to substation equipment image processing, a method that performs feature extraction for recognition through substation equipment images. The research focuses on the expansion of the image sample set, the automatic training method based on recognition rate, and the voting strategy based on integrated learning, which not only improves the training efficiency of the model but also increases the recognition rate, and the proposed method is of high practicality.

## 1. Introduction

Since the 20th century, computer technology has contributed to a great improvement in the way all sectors of society produce. The traditional way of manually manning substations has been made possible by the continuous advancement of automation technology, which has made it possible for unmanned substations, and through which companies hope to change the current backward status quo of substations. In this social environment, video monitoring systems have received widespread attention, and more and more people are devoted to the study of video monitoring, which has become a hot research topic today.

Substation video surveillance systems need to detect and identify specific targets in the surveillance video taken by inspection robots [1], and the analysis of specific targets to predict and understand the significance of their behavior. At its core, it analyses the behavior of targets and events by processing the original video image through a series of

algorithms such as feature extraction [2], image segmentation, image target detection and tracking, and image classification and recognition. In order to improve these hardships, we need to improve the operational efficiency and quality of substation equipment. Robot-based automatic inspections are implemented to fix the video images collected by robots. Tracking and identification can then understand the operating status of each equipment in the substation and improve the safety and reliability of the operation and maintenance of the substation.

Existing video image target detection and tracking algorithms have a low accuracy rate due to poor image quality from a distance, the complexity of the outdoor environment, and the presence of multiple interferences [3]. Therefore, the existing video surveillance system cannot be effective for video content analysis, understanding, processing, providing solutions to complex power problems, and meeting the security requirements of today's unattended substations. In summary, the power

system is in urgent need of basic and applied theories based on intelligent image tracking and image recognition of substation equipment.

Power equipment condition detection includes infrared detection technology, dissolved gas analysis in oil, and partial discharge detection methods. A large number of empirical studies have shown that most electrical equipment will show temperature changes when faults occur [4]. Infrared inspection technology is the only way to reveal the operating status of power equipment through visualization of equipment temperature information and can be used for nonstop, regular, or real-time inspection of power equipment, with noncontact, fast, and safe features. The current substation equipment infrared image analysis is still mainly manual, the influence of human factors in this method is large, and there is the problem of low efficiency of image analysis. Due to the limitations of infrared detection technology itself and the complexity of power equipment, the operation status of power equipment can no longer be judged by infrared detection technology alone. With the development of artificial intelligence and computer image recognition technology, the combination of substation equipment and image processing technology has become a new breakthrough [5].

The field of artificial intelligence has evolved rapidly over the past decade, with deep learning [6] (DL) being one of the fastest growing machine learning methods. DL is formed by using multiple simple neurons to form a multilayer network and adjusting the input according to the nonlinear relationship between the input and output. DL is essentially a complex function that extracts features [7] from an input sample and reflects its power. In recent years, DL has led to many research results in areas such as video tracking and video recognition [8]. When DL is integrated into the remote video monitoring system of substations to automatically track and identify the status of video monitoring equipment, it can greatly improve the accuracy of automatic equipment status identification, realize the automatic detection function of substations, ensure the operation of substations, and improve maintenance efficiency. On the other hand, in the complex outdoor environment, various interferences may occur, resulting in blurred and otherwise inaccurate video from the substation. Deep learning technology has strong feature extraction ability in image recognition and has strong generalisation ability. Therefore, in this study, DL was combined with image processing technology and category detection and pointer reading of substation images was introduced. The theoretical and applied research on image tracking in substations and image recognition based on DL and image feature extraction is of great research importance and has many potential applications.

As early as 1989, LeCun et al. published a study on CNNs and named the structure LeNet5 [9], which has a very high recognition rate for image figures. However, due to the small number of layers, this model structure did not show good recognition results for image data with many image features. However, with the improvement of computer hardware, multilayer CNNs and DL have become a current research hotspot. The network structure has greatly reduced the complexity of the network structure and reduced the

requirements for hardware, making the network model accessible to an increasing number of people and facilitating its widespread dissemination. Multilevel CNN model can use images directly as outputs, avoiding the complex process of feature extraction [10] and data reconstruction found in image processing algorithms. CNN is a multilayer perceptron specially designed for the recognition of two-dimensional shapes and has highly undistorted characteristics for the translation, rotation, scaling, and offset of objects in images.

DL is a CNN with hidden layers. In supervised learning, as long as the sample data are manually labelled [11, 12], the deep CNN can automatically extract features from a specific target in the substation image according to the labelling, recognise, analyse, and understand the target, achieve remote intelligent inspection instead of manual inspection to transcribe meter values, analyse the status of a specific monitoring object on-site, and upload the results of the analysis (standard information, pictures, or video information). The results of the analysis (standard information, pictures, or video information) are uploaded to the unified video monitoring platform of the power grid to meet the requirements for monitoring the safety and production of individual meters in smart substations.

The use of CNN methods for target identification is subject to interference from many sources, and the inaccuracy of one parameter may lead to a significant reduction in the final recognition rate, with too many uncontrollable factors. Therefore, we can consider combining CNN methods with other machine learning algorithms, reducing the requirement for neural network tuning, allowing for a certain amount of error in the neural network, and by combining several algorithms, even higher recognition results can be achieved. In this study, CNNs are introduced to the processing of substation equipment images, combining them with traditional image processing methods for feature extraction, classification, and recognition of substation equipment images. The research focuses on the expansion of the image sample set, the automatic training method based on recognition rate, and the voting strategy based on integrated learning, which not only improves the training efficiency of the model but also increases the recognition rate, and the proposed method is of high practicality.

## 2. CNN Models

*2.1. Model Introduction.* As an important branch of machine learning, DL has broken the bottleneck in the development of artificial intelligence based on the construction of multi-implicit layer machine learning models and learning more effective features through massive data and has promoted technological development in many fields. The idea of DL was first introduced by Geoffrey Hinton [13].

In the 1980s, the LeNet5 network model was proposed. Due to the limitations of the computer hardware level at that time, the number of layers of the LeNet5 model was limited. When working with complex image recognition, the recognition effect of the model is not good. With the computer hardware available at the time, algorithms such as support

vector machines (SVMs) performed better than CNNs. However, shallow machine learning algorithms have their limitations, as they have fewer parameters than DL, and this has led to a major bottleneck in the accuracy of the shallow algorithms, which has prevented them from being used on a large scale.

However, technology has always been developing, and after entering the 21st century, computer hardware has been developed rapidly, with the emergence of various high-performance CPUs and GPUs, which has paved the foundation for the development of DL. High-performance computer hardware solves the problem of computation, and excellent CNN algorithms solve the problem of algorithms. As a result, CNNs have grown exponentially. The number of network layers is also increasing, successfully achieving better and better results in image recognition, speech recognition, financial prediction, and judgement. In the 2012 ImageNet competition, deep CNNs were a hit in the field of image recognition, surpassing the SVM algorithm by a huge margin and attracting great attention from scholars at home and abroad.

CNNs [14, 15] are an important algorithm in DL, whose main features are reflected in three aspects: (1) having local awareness, using local connections between adjacent neurons instead of full connections, i.e., having convolutional layers that perform convolutional operations with the data. (2) Using weight sharing and greatly reducing the parameters required for training. (3) Using pooling operations to achieve dimensionality reduction and reduce the occurrence of overfitting. This makes the network more tractable and robust. Based on the above characteristics, CNNs have gained extensive attention, research, and application in image classification problems. LeNet-5 CNN has an excellent performance in classifying MNIST handwritten digital dataset [16, 17], which opened a new phase of CNN in the field of image classification, and since then, AlexNet [18], VGGNet [19], GoogleNet [20] and other CNNs have been proposed, which have made great progress in image classification and recognition.

## 2.2. Model Structure

**2.2.1. The LetNet-5 Model.** The convolutional neural network is mainly composed of an input layer, a convolutional layer, a pooling layer, a fully connected layer, and an output layer. The neural network is to alternately connect several convolutional layers and pooling layers, that is, a convolutional layer is connected to a pooling layer, and the pooling layer is connected to a convolutional layer. The network structure of the convolutional neural network model LetNet-5 is shown in Figure 1.

The convolutional layer is the core component of a CNN and consists of a series of convolutional kernels. The convolutional operation is as follows:

$$C_{p,j}^{(l)} = f \left( \sum_{l \in M_j^{(l-1)}} \sum_{(u,v) \in k^{(l)}} \left( W_{ij(u,v)}^{(l)} \otimes X_{p,j}^{(l-1)}(x+u, y+v) + b_j^{(l)} \right) \right), \quad (1)$$

where  $C_{p,j}^{(l)}$  is the feature image prime value of the  $p$  training sample in layer  $l$ .  $W_{ij(u,v)}^{(l)}$  denotes the convolution kernel in layer  $l$ ,  $u$  and  $v$  are its step size,  $b_j^{(l)}$  is the bias,  $X_{p,j}^{(l-1)}$  is the feature map in layer  $l-1$ ,  $M_j^{(l-1)}$  is the input mapping in layer  $l-1$ , and  $l$  is the current layer activation function.

The output of the convolutional needs to be nonlinearly mapped by excitation functions such as sigmoid, tanh, and ReLU. The sigmoid and tanh functions are similar and are expressed as follows:

$$\begin{aligned} f(x) &= \frac{1}{1 + e^{-x}}, \\ \tanh(x) &= \frac{2}{1 + e^{-2x}} - 1. \end{aligned} \quad (2)$$

The images of these two functions are shown in Figure 2.

As can be seen in Figure 2, the tanh function converges more quickly than the sigmoid function. Both activation functions suffer from soft saturation, i.e., when the input falls into the saturation zone, it tends to cause the gradient to disappear, thus making the model training slower. The ReLU function is more expressive than the above activation functions, and the gradient of the non-negative interval of ReLU is constant, which can overcome the problem of gradient disappearance, thus maintaining the convergence speed of the model in a stable state. Therefore, the ReLU excitation function is widely used in CNNs with the expression  $f(x) = \max(0, x)$  and its image is shown in Figure 3.

The pooling layer is located in the middle of successive convolutional layers and its main role is to reduce the amount of image data, while ensuring that local features do not change. It speeds up CNN training and also prevents overfitting. Max pooling calculates the maximum value of a local region in the feature map and uses this value as the pooled result. Average pooling is the calculation of the average of the image regions and takes this as the resultant value.

The CNN training has two stages: the forward propagation of the signal and the backward propagation of the error. First, the network weights are initialised and the input infrared image data are passed through convolutional, pooling, and fully connected layers to obtain the output value. The backward propagation refers to the construction of the error function, when the degree of error between the output value and the target value exceeds the set threshold, the error is propagated backwards into the network, and the network weights parameters are adjusted according to the loss function and gradient descent method for another training until the end of training, and its training process is shown in Figure 4.

**2.2.2. Underfitting and Overfitting.** Fitting is the process of taking some discrete data and adjusting the parameters of the function through an algorithm to minimise the sum of the distance between the fitted function value and the data points. The resultant function has the smallest distance between the point set and these discrete data. The effect of underfitting is that the recognition effect of the trained

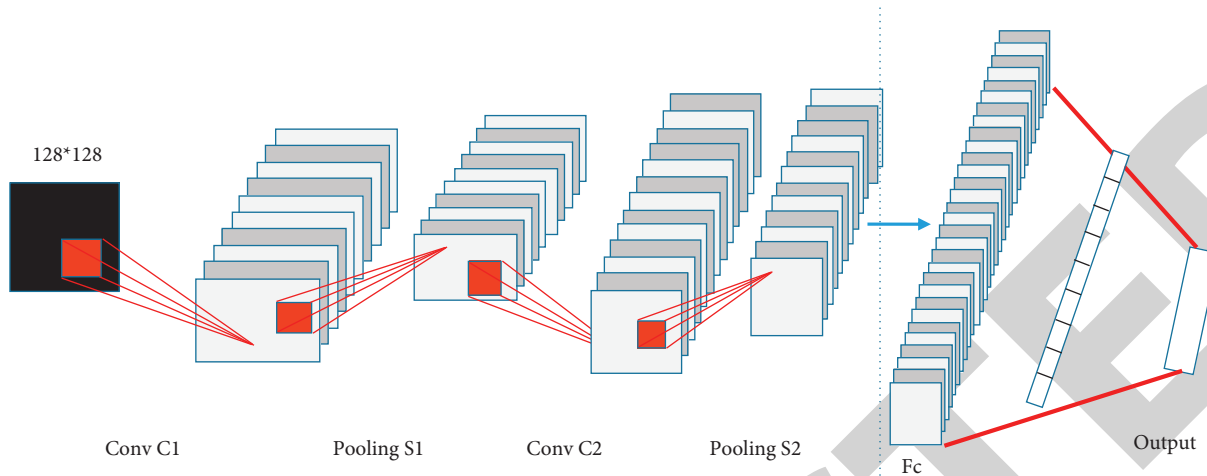


FIGURE 1: CNN structure.

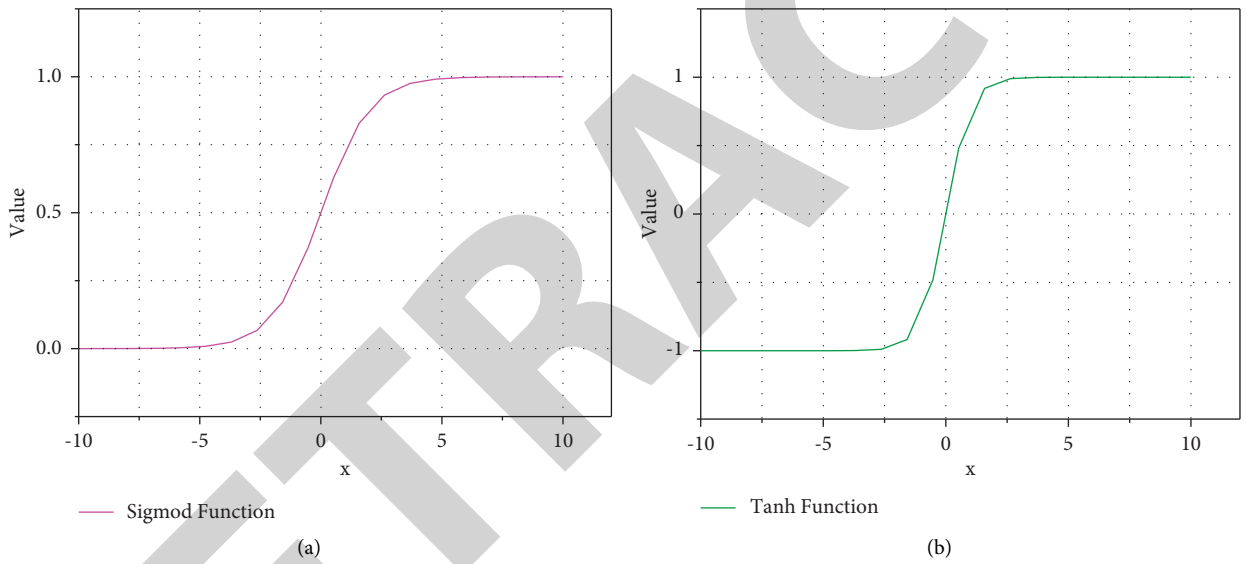


FIGURE 2: Images of sigmod and tanh functions. (a) Sigmod. (b) Tanh.

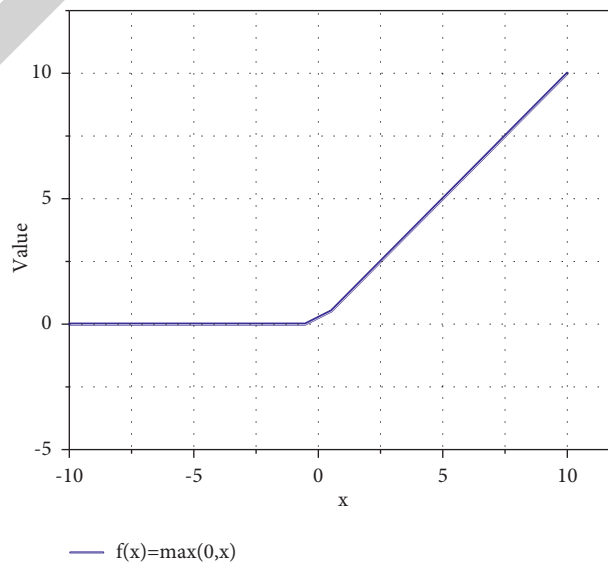


FIGURE 3: ReLU excitation function image.

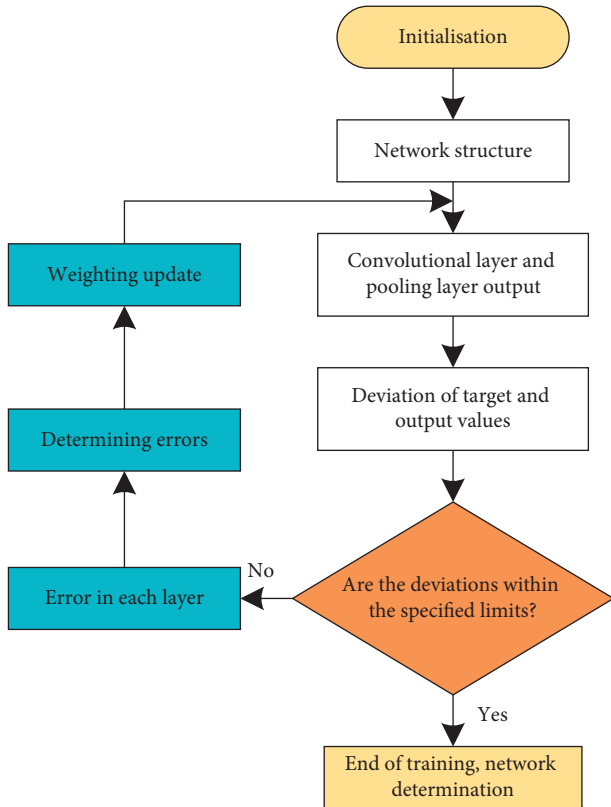


FIGURE 4: Schematic diagram of the CNN training process.

model in the training set and the test set is not good. There are two methods to solve the problem of underfitting, one is to increase the amount of data, and the other is to reduce the complexity of the designed neural network.

Overfitting is manifested by the fact that as the number of training sessions increases, the trained model performs very well in the training set but flattens out in the test set. The reason for this is the inappropriateness of supervised learning, where the backpropagation algorithm iterates backwards to modify the parameters and learn features such as noise as if they were correct. However, the effect of these noisy features should actually be reduced.

### 3. Data Processing

Most of the traditional instrumentation type recognition projects are pointer recognition for a certain type of instrumentation community, or require manual selection of a certain type of instrumentation and then inputting image data for reading. But for intelligent substations and inspection robots, this type of method is not applicable, the inspection robot cannot automatically get the type of instrumentation equipment, and manual operation is too cumbersome. Therefore, we need the inspection robot to “learn” how to recognise which type of instrument the image data collected is, so that the image data can be recognised using the appropriate template and data.

In the present day, computer hardware is developing very fast and is sufficient to support the computational load of deep neural networks in multiple layers. In recent years,

the results of DL in the field of image recognition are well known, so in this study, we will use deep neural networks to identify the type of substation equipment from the image data collected by the inspection robot.

**3.1. Data Set Expansion.** Deep neural networks have a strong fitting capability due to their deeper neural network layers. If the training set is small, it will lead to an overfitting condition, which will inevitably lead to weak generalisation. The results are often unsatisfactory when testing instrumentation data that is in outdoors and a highly variable environment.

However, in the case of outdoor smart substations, there are safety hazards or environmental factors that prevent us from obtaining sufficient data sets. To solve the problem of inadequate data sets, this study uses background replacement, rotation, affine, translation, and noise addition to expand the image data set.

**3.1.1. Image Background Replacement.** The backgrounds and environments, in which instrumentation devices are located, vary little. In order to achieve the diversity and complexity of the image background, the replacement of the background of the instrumentation equipment is achieved by means of image processing. The image augmentation method of background replacement is used for several reasons:

- (1) The background replacement requires a large amount of calculation and has a high success rate, which can lead to a high time complexity if the initial trial data volume is large.
- (2) After the background replacement, the instrument is placed at randomly chosen coordinates, and a richer amount of data can be obtained.
- (3) The manual workload is greatly reduced due to the smaller amount of data and the manual screening of the meter scene replacement if errors occur.

Figure 5 shows the replacement effect after the resolution of the background image was modified to 1000 \* 1000.

**3.1.2. Image Panning and Expansion.** The position, in which the meter is located in the data captured by the inspection robot, is uncertain. When we manually capture image data, we often shoot the meter data in the centre of the image, which is not desirable for later generalisation to identify the input data of the inspection robot. We therefore need to manually perform a translation transform for the dataset so that the meters will exist anywhere in the image, not only expanding the dataset but also enabling the final neural network model to be generalised much more.

The translation transform is a simple transformation where all points on an image are moved horizontally along the  $x$  and vertically along the  $y$  according to a given offset, and is given by

$$\begin{cases} x = x_0 + \Delta x, \\ y = y_0 + \Delta y. \end{cases} \quad (3)$$

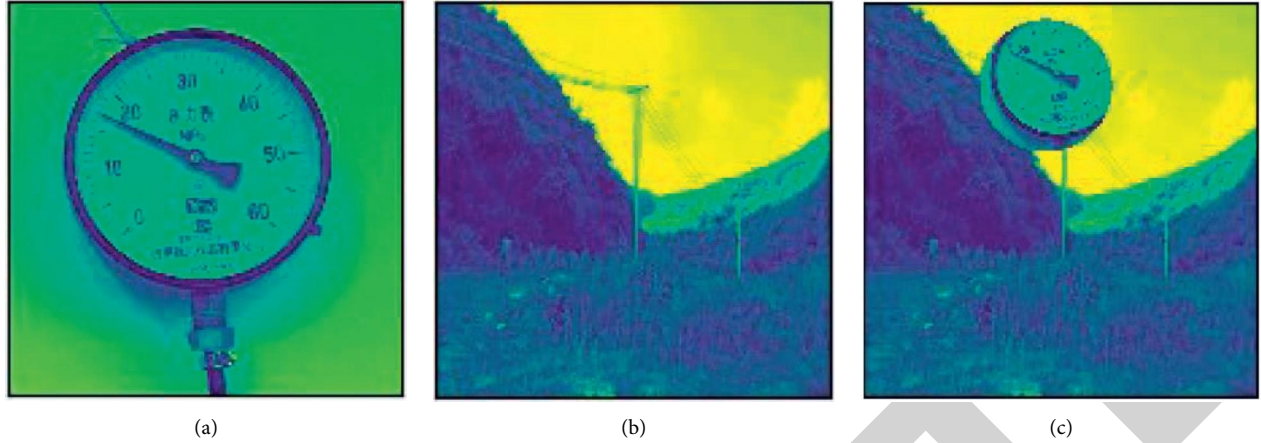


FIGURE 5: Image background substitution effect. (a) Original image, (b) background image, and (c) background replacement.

The equation can be expressed in terms of matrix transformations:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \Delta x \\ 0 & 1 & \Delta y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix}. \quad (4)$$

After panning, the choice was made to use white to fill in the excess panned out in order to reduce the distracting elements in the image; in practice, black would cause a greater amount of noise.

**3.1.3. Image Rotation Expansion.** The instrument can be panned to one side by a translation transformation, but there is no guarantee that the instrument will be at a particular position in the graph when the inspection robot is actually collecting data. If the devices in the image dataset used are all at a particular location, then the deep neural network model trained using this dataset will not be available for some of the images acquired by the inspection robot. The CNN model may not achieve accurate recognition, i.e., the depth model does not have excellent generalisation capabilities. In order to make the training model have better generalisation ability and solve the impact of insufficient data sets on model training, image rotation and translation methods are used to analyse the data. At the same time, it can increase the diversity of samples and effectively increase the generalisation ability of training results.

Each image is rotated and transformed to expand the dataset by a factor of eight, which greatly expands the dataset, and after translation and rotation, the positions of the substation images are present at various locations in the image, increasing the diversity of the dataset.

## 4. Experimental Results and Analysis

### 4.1. Experimental Preparation

**4.1.1. Network Structure and Parameter Setting.** The issue regarding the choice of convolutional kernel size in each convolutional layer is different for different classes of

substation equipment images. A large convolutional kernel can easily extract the overall features of the instrument but ignore some detailed features, while a small convolutional kernel, in contrast, can extract detailed feature information but may ignore the overall information of the instrument. The main problem in the structural design is the confirmation of the hidden layer structure, and the relevant parameters designed in this process include the convolution kernel. The size of the feature maps, the size of the pooling kernel, the number of iterations, and the number of batches are influencing factors.

The pixel values of the normally acquired image data are relatively high, but for recognition, such high precision image data are not required and are very consuming in terms of training difficulty and time spent. Therefore, if the image data are uniformly sized at  $128 * 128$  when inputting the acquired data, the main features will be preserved and the training speed will be significantly increased.

**4.1.2. Selection of the Number of Feature Maps.** The number of feature maps in the hidden layer  $n$  means that the image uses  $n$  different  $N * N$  convolution kernels during the convolution machine operation or  $nM * M$  pooling kernels during pooling. In general, the number of feature maps is chosen based on personal experience, but for those using the software, it is not possible to set the number of different layers for each type of image data through experience. Therefore, in this study, the number of feature maps in the convolution and pooling layers are selected automatically from the experimental results, thus avoiding the complicated operation of the staff and improving the recognition rate reduction due to inexperience.

**4.1.3. Initialisation of Weights.** The initialization operations are first performed on the weight matrices in the convolutional and fully connected layers. The multiple iterations of a deep neural network are designed to continuously modify the values in this weights matrix, and therefore, the initialization of weights affects the final recognition rate. If

the range of variation in the initialization of the weights is too large, it will result in more iterations being required to achieve optimal recognition. In this study, the set range of the aggregated weights is  $[-1, 1]$ , and all values in the weight matrix of  $N * N$  are taken randomly from this range. Because the range is small, the number of iterations required is not too large, which greatly improves the speed of computation and achieves the “optimal solution” as quickly as possible.

**4.1.4. Selection of the Activation Function.** The activation function used in this study is ReLU, which is a modified version of ReLU, and is used to solve the problem of necrosis. The main difference with ReLU is that instead of having a value of 0 when  $x < 0$ , it becomes  $ax$ , where  $a$  is a smaller value, in this study  $a = 0.15$ .

$$ReLU(x) = \begin{cases} x, & x > 0, \\ ax, & x \leq 0. \end{cases} \quad (5)$$

**4.2. Parameter Settings.** The parameter settings of the CNN in the study are listed in Table 1.

For the above parameters, the model is randomly reset each time it is executed, and the parameters are saved. If the result of this training does not reach the set recognition rate, it will be reset once again and with different parameters than previously saved until the recognition results reach the set threshold. The neural network model can therefore be applied to the training of any type of substation image data.

**4.3. Experimental Results.** The experimental dataset consisted of 8000 image data and a total of 47 runs of the neural network structure with different parameters.

The relationship between the recognition rate of the CNN model used and the number of different iterations is listed in Table 2.

A visual comparison of the recognition rate of the CNN for different numbers of iterations of the substation equipment image is shown in Figure 6. In Figure 6, the best recognition rate of the substation equipment image is achieved when the number of iterations is 8.

In this training process, the recognition rate of the training set and the recognition rate of the test set are listed in Table 3.

A comparison of the recognition rate data between the training set and the test set for nine different iterations of the substation equipment images is shown in Figure 7.

It can be seen that at iteration number 1, the recognition rate of the image is too low and the model is not performing as it should at this point and there is no value in the study at this point. The recognition time data for single images with iterations 5–9 are listed in Table 4.

A visual comparison of the recognition times of the training and test sets for a single image at different numbers of iterations is shown in Figure 8.

TABLE 1: Parameter settings of the CNN.

Weighting	Normal distribution initialization [0, 1]
Learning_rule	Sad
Learning_rate	0.015
Learning_momentum	0.9
Weight_decay	0.001
n_iter	1–9
n_stable	10
f_stable	0.001
valid_size	0.1
Verbose	True

TABLE 2: Recognition rate data for different number of iterations.

Number of iterations	Recognition rate (%)
1	93.6
2	97.1
3	97.2
4	97.9
5	98
6	98
7	96.9
8	98.57
9	98.5

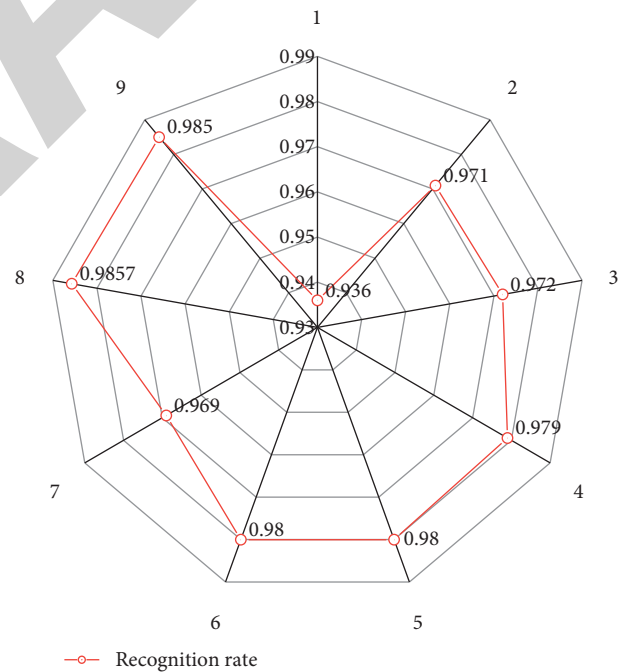


FIGURE 6: Visual comparison of recognition rates of substation equipment images with different iterations.

#### 4.4. Result Analysis and Recognition Rate Optimization.

The maximum number of iterations was set to 9. However, in some structures, 9 iterations did not reach the optimal number of iterations for the structure, and the training time increased significantly as the number of iterations increased. The model running time and the model recognition accuracy are comprehensively considered. When the number of iterations is set to 9, the recognition accuracy of the model is



TABLE 3: Recognition rate data of the training set and test set for different number of iterations of substation equipment images.

Number of iterations	Train score (%)	Test score (%)
1	93.5	94.5
2	97.2	98.1
3	97.3	97.2
4	97.6	97.5
5	98.1	97.2
6	97.9	97.4
7	96.7	96.2
8	98.6	97.8
9	98.3	97.4

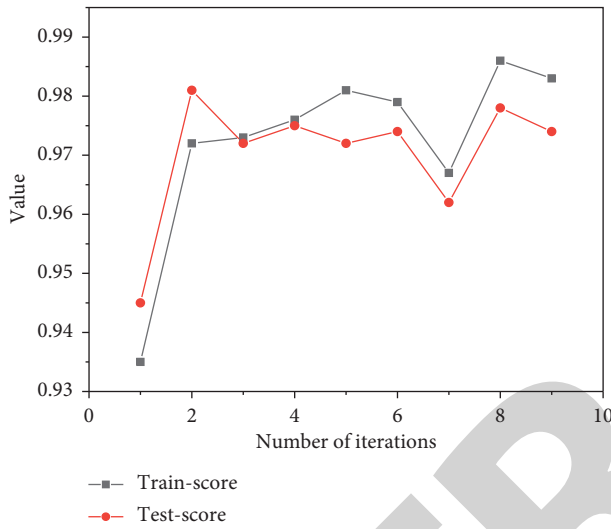


FIGURE 7: Comparison of recognition rates between training and test sets with different iterations.

TABLE 4: Comparison of recognition times (s) for individual images.

Number of iterations	Train score	Test score
5	0.24	0.21
6	0.26	0.23
7	0.27	0.24
8	0.29	0.25
9	0.30	0.27

the best and the running time at this time has less impact on the model. The main idea of integrated learning [21, 22] is to build three neural network structures with high performance and diversity and to combine the three learners to recognise images through a voting strategy. High performance means that each neural network has a relatively high recognition rate, without using multiple “weak learners” for learning.

The voting strategy is often used for prediction of classification problems, and the core idea is that the few follow the many. The prediction category is assumed to have  $N$  classes, and after using multiple neural network models with different parameters to make predictions, the final classification category is based on the one with the highest number of identified classes.

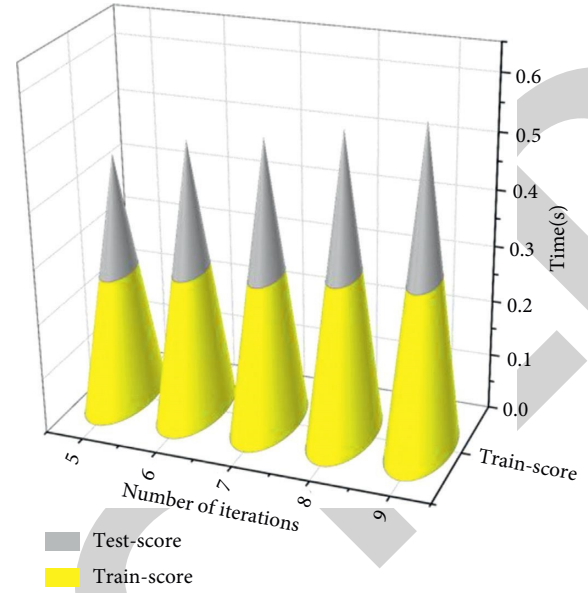


FIGURE 8: Comparison of recognition time for a single image with different number of iterations.

TABLE 5: Five times recognition results of the three neural network models.

Category	Model 1 (%)	Model 2 (%)	Model 3 (%)
1	98.7	98.6	98.4
2	98.6	98.7	98.5
3	98.8	98.5	98.3
4	98.7	98.5	98.3
5	98.7	98.7	98.5

After training the neural network 47 times with different structures, the three neural network models with the highest recognition rates were selected among them and their recognition data are listed in Table 5. Each neural network was experimented five times and the average value was selected as the final result.

The five recognition results for the three neural network models are shown in Figure 9.

The average recognition rates of the three models were 98.7%, 98.6%, and 98.4%, respectively. Using the voting method strategy to combine the three results, the recognition rate data under five experiments are listed in Table 6. Where time conditions allow and the computer configuration is high, an attempt to increase the upper limit of the number of iterations should result in even higher recognition rates. Integrated learning refers to the construction and combination of multiple learners to complete a learning task, sometimes referred to as multiclassifier systems, committee-based learning, etc., and can often achieve significantly better generalisation performance than a single learner. This method is particularly effective for “weak learners.”

As seen in Table 6, the most this recognition rate reached 99.4%. As the computer configuration increases, the upper limit of the number of iterations is also increased, and the use of the voting method strategy further improves the

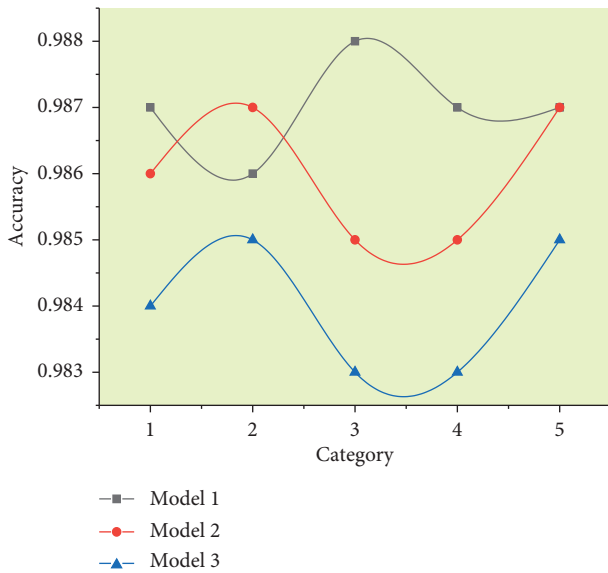


FIGURE 9: Comparison of the five recognition results of the three neural network models.

TABLE 6: Recognition rate data for the voting method strategy for five experiments.

Category	Accuracy (%)
1	99.3
2	99.4
3	99.4
4	99.3
5	99.4

recognition accuracy of the substation equipment. The recognition time for a single image can be controlled compared to previous algorithms, demonstrating the effectiveness of the proposed method.

## 5. Conclusion

Due to the complexity of the current substation environment, the variety of existing equipment and the presence of multiple interferences make the accuracy of image processing algorithms low. This study applies convolutional neural networks to the analysis of substation equipment images, using image translation and rotation methods to expand the sample set and replace the image background. Experiments are carried out on substation equipment images acquired by inspection robots, using automatic training methods for neural networks and an integrated learning voting strategy for substation equipment image recognition. The proposed method improves the training efficiency of the model and achieves a higher recognition rate by combining three strong learners through the voting rule, with an optimal substation equipment image recognition rate of 99.4%. The proposed method is ideal for substation image recognition experiments, but there are more factors that may interfere with the recognition of real devices.

## Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] M. Yang, S. Ji, W. Xu et al., "Detecting Human Actions in Surveillance Videos," vol. 65, no. 3, pp. 128–132, 2011.
- [2] A. Rehman and T. Saba, "Features extraction for soccer video semantic analysis: current achievements and remaining issues," *Artificial Intelligence Review*, vol. 41, no. 3, pp. 451–461, 2014.
- [3] W. Hu, *Research on Target Detection and Tracking Algorithms Based on Video Images*, Changjiang University, Jingzhou, Hubei, China, 2017.
- [4] H. Y. Zhang, *Research on Some Key Technologies for Early Warning and Prediction of Current-Carrying Faults in Power Equipment*, Zhejiang University, Hangzhou, Zhejiang, China, 2014.
- [5] H. Zhuang li, H. Tong, and Z. Yihui, "Fast image recognition of transmission tower based on big data," *Protection and Control of Modern Power Systems*, vol. 3, no. 3, pp. 149–158, 2018.
- [6] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [7] M. Ranzato and Y. LeCun, "A sparse and locally shift invariant feature extractor applied to document images document Analysis and Recognition," vol. 2, pp. 1213–1217, in *Proceedings of the ICDAR 2007. Ninth International Conference on*, vol. 2, pp. 1213–1217, IEEE, Curitiba, Brazil, September 2007.
- [8] Y. Pan, T. Mei, T. Yao, and Y. Rui, "Jointly modeling embedding and translation to bridge video and language," *Computer Science*, vol. 3, no. 6, pp. 4594–4602, 2015.
- [9] Y. Lecun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Proceedings of the 2010 IEEE International Symposium on Circuits and Systems*, vol. 14, no. 5, pp. 253–256, IEEE, Paris, France, May 2010.
- [10] D. J. Hurley, M. S. Nixon, and J. N. Carter, "Force field feature extraction for ear biometrics," *Computer Vision and Image Understanding*, vol. 98, no. 3, pp. 491–512, 2005.
- [11] C. Quan and F. Ren, "Unsupervised product feature extraction for feature-oriented opinion determination," *Information Sciences*, vol. 272, no. 3, pp. 16–28, 2014.
- [12] D. Erhan, Y. Bengio, A. Courville, P. A. Manzagol, and P. Vincen, "Why does unsupervised pre-training help DL?" *Journal of Machine Learning Research*, vol. 11, no. 3, pp. 625–660, 2010.
- [13] G. Hinton, S. Osindero, M. Welling, and Y. W. Teh, "Unsupervised discovery of nonlinear structure using contrastive backpropagation," *Cognitive Science*, vol. 30, no. 4, pp. 725–731, 2006.
- [14] B. Chen, H. Li, W. Luo, and J. Huang, "Image processing operations identification via CNN," *Science China Information Sciences*, vol. 63, no. 3, pp. 31–45, 2020.
- [15] L. Zhang, Z. Sheng, Y. Li, Q. Sun, Y. Zhao, and D. Feng, "Image object detection and semantic segmentation based on

- convolutional neural network,” *Neural Computing & Applications*, vol. 32, no. 7, pp. 1949–1958, 2019.
- [16] Y. Pei and L. Ye, “Cluster analysis of MNIST data set,” *Journal of Physics: Conference Series*, vol. 2181, no. 1, pp. 45–49, 2022.
- [17] J. Liu and C. Fu, “MNIST data set recognition research based on TensorFlow framework,” *International Core Journal of Engineering*, vol. 7, no. 2, pp. 410–414, 2021.
- [18] Y. Hendrawan, “AlexNet CNN to classify the types of Indonesian coffee beans,” *IOP Conference Series: Earth and Environmental Science*, vol. 905, no. 1, pp. 691–695, 2021.
- [19] H. Ke, D. Chen, X. Li, Y. Tang, T. Shah, and R. Ranjan, “Towards brain big data classification: epileptic EEG identification with a lightweight VGGNet on global MIC,” *IEEE Access*, vol. 6, no. 9, Article ID 14722, 2018.
- [20] Y. Fu, “Circular fruit and vegetable classification based on optimized GoogLeNet,” *IEEE Access*, vol. 9, no. 3, Article ID 113599, 2021.
- [21] Y. Chen, C. Dong, and B. Wu, “Crown profile modeling and prediction based on ensemble learning,” *Forests*, vol. 13, no. 3, pp. 410–411, 2022.
- [22] J. Zuo, W. Zhao, L. Chen et al., “Multi-focus image fusion algorithm based on random features embedding and ensemble learning,” *Optics Express*, vol. 30, no. 5, pp. 8234–8247, 2022.