

Enhancing Automation with Label Defect Detection and Content Parsing Algorithms

Min Zheng

Institute of Aeronautical Engineering, Ezhou Vocational University, Ezhou, Hubei, China

The stable operation of power transmission and distribution is closely related to the overall performance and construction quality of circuit breakers. Focusing on circuit breakers as the research subject, we propose a machine vision method for automated defect detection, which can be applied in intelligent robots to improve detection efficiency, reduce costs, and address the issues related to performance and assembly quality. Based on the LeNet-5 convolutional neural network, a method for the detection of character defects on labels is proposed. This method is then combined with squeezing and excitation networks to achieve more precise classification with a feature graph mechanism. The experimental results show the accuracy of the LeNet-CB model can reach up to 99.75%, while the average time for single character detection is 17.9 milliseconds. Although the LeNet-SE model demonstrates certain limitations in handling some easily confused characters, it maintains an average accuracy of 98.95%. Through further optimization, a label content detection method based on the LSTM framework is constructed, with an average accuracy of 99.57%, and an average detection time of 84 milliseconds. Overall, the system meets the detection accuracy requirements and delivers a rapid response, making the results of this research a meaningful contribution to the practical foundation for ongoing improvements in robot intelligence and machine vision.

ACM CCS (2012) Classification: Computing methodologies → Machine learning → Machine learning algorithms

Computing methodologies → Artificial intelligence → Computer vision

Keywords: machine vision, label characters, detection, LeNet-5, neural network, LSTM

1. Introduction

With the development of industry, effective use of electricity has become the core of sustainable development for enterprises. In this context, the low-voltage circuit breaker, as the core equipment in the power system, is experiencing an increasing demand. To ensure high efficiency and product quality in the production process, it is particularly important to automate the appearance and assembly quality of the low-voltage circuit breaker. However, the current detection process mainly relies on manual operation. This method has many limitations, such as strong subjectivity, low efficiency, and poor stability, all of which significantly impact the overall detection effectiveness and production efficiency [1–3].

In light of these challenges, the research aims to overcome the limitations of existing methods and address these issues by enhancing the automation and intelligence of detection processes. Two key algorithms, label defect detection and content resolution, are developed. The goal of these algorithms is to improve the label defect detection accuracy and clarity of the character electrical parameter identification, thereby solving the common misidentification and confusion in manual detection problems [4–6].

In the technical implementation, the LeNet-5 Convolutional Neural Network (LeNet-5) is adopted. Subsequently, LeNet-5 is combined with the Squeeze Excitation Network (SENet) to develop an advanced label feature detection

method based on the feature graph channel mechanism. In addition, a label content parsing scheme based on a Long Short-Term Memory (LSTM) network is constructed. This scheme effectively improves classification performance and better meets the needs of intelligent robots in complex industrial environments.

The rest of this paper is organized as follows. Section 2 summarizes the research achievements and shortcomings of machine vision and its application in detection of character defects on labels. In Section 3, a method based on a convolutional neural network model is proposed for content parsing and detection of faulty characters in labels. The experiments conducted on the designed model are described in Section 4. Section 5 concludes the paper by providing a summary of the experimental results and the shortcomings in the research and proposes future research directions.

2. Related Work

Machine vision offers unique advantages in character defect recognition, with deep learning and neural network technology playing pivotal roles [7]. Both industry experts and scholars have conducted a series of related studies in this field.

Li *et al.* proposed a hybrid model for optimizing code smell detection through multi-level code representation. The model first parsed the code into an abstract syntax tree, and a Convolutional Neural Network was used to perform grammar and semantic prediction. A bidirectional LSTM network with an attention mechanism was used to analyze encoded symbols. The experimental results showed that the model outperformed existing methods in code smell detection for both single-label and multi-label codes [8].

W. Tang *et al.* proposed an evolutionary algorithm that combined traditional image processing, deep learning, and transfer learning to achieve automatic recognition of unknown defects. This algorithm extracted and classified features through deep learning, using deep clustering to classify and store unmarked defects, and automatically update the database. Then, transfer learning was introduced to train classifiers using updated databases. The effective-

ness and accuracy of this method in image defect detection were verified [9].

C. Zheng *et al.* used label decoupling technology to split the original label map into internal and boundary diffusion maps to collaborate with GT maps. This process involved a three-stream neural network with a multi-scale interactive expansion module designed to explore more contextual information. The boundary perception feature stitching modules based on attention were developed to integrate multi-modal information. This strategy significantly improved detection accuracy and boundary clarity [10].

Potapenko *et al.* explored the accurate detection of retinal edema using convolutional neural networks on optical coherence tomography images. Training data, with recorded treatment decisions as labels, effectively identified the presence of edema, maintaining high accuracy even in the presence of inconsistencies between training and validation labels [11].

Kumar *et al.* proposed a multi-label learning algorithm. The dataset was divided into three groups by gender. The word frequency, inverse document frequency, and global vector features were calculated for each group. The analysis involved the problem transformation method and a multi-channel recurrent neural network with attention mechanisms. The results showed that traditional multi-label conversion methods had excellent processing performance for small amounts of data and long sequences [12].

Wang *et al.* proposed a deep learning framework for detecting related events from news reporting on renewable energy system accidents. This framework adopted bidirectional encoder representation and text Convolutional neural network. The framework trained a binary classifier for detecting event existence and a multi-label classifier for identifying event attributes. The experimental results showed that the framework performed well in detecting the existence of events [13].

Turkoglu *et al.* proposed a deep neural network method based on an Extreme Learning Machine (ELM). Convolutional neural networks were used to extract features from scanned images. This method utilized a pre-trained deep network architecture and transfer learning to evaluate the performance of ELM classifiers with different

activation functions for effective detection and recognition. The experimental results demonstrated the effectiveness of the model [14].

Zhou *et al.* proposed an optimized fault detection method using a deep learning model. Transfer learning was used to mitigate the impact of data differences while learning fault-related features. The model also explored the common features of real and synthetic seismic data to enhance their applicability. The experimental results showed that this method significantly improved detection performance and reduced manual labeling costs without the real data labels [15].

In summary, many scholars and scientists have dedicated efforts to improve detection accuracy and apply it across diverse fields. However, considering the significant diversity, size differences, and irregular morphology of characters on labels, a LeNet-5 neural network is proposed for content parsing and detection of defects in characters on labels.

3. Research Model

Considering the diversity, size differences, and irregular shapes of label characters, we have developed a character defect detection method based on neural networks, namely the LeNet-5 architecture. This method incorporates a feature graph channel mechanism and is designed by combining LeNet-5 and SENet. To further enhance classification performance and meet the demands of intelligent robots, a label content parsing scheme based on LSTM is constructed.

3.1. Character Defect Detection Model Based on Convolutional Neural Networks

Research on machine vision detection algorithms for robot intelligence is a pivotal area that can enable robots to accurately recognize and understand the visual information from their surroundings, and achieve efficient and intelligent operations. The label characters exhibit a variety of fracture defects, such as top and bottom, lateral, combination, and par-

tial fracture types. There are differences in the morphology and size of these defects. Therefore, relying on traditional feature engineering detection methods makes it difficult to achieve stable reliability.

To address this challenge, a novel machine vision detection algorithm is proposed to accurately identify and classify these fracture defects. The LeNet-5 detection algorithm, as the core tool used in this study, is mainly used for precise recognition of label character defect detection methods [16].

Figure 1 shows the LeNet-5 network architecture. It includes three convolutional layers for feature extraction, two pooling layers for downsampling, and a fully connected layer for feature mapping, as well as input and output layers. LeNet-5 integrates feature extraction and classification, laying the foundation for Convolutional Neural Networks [17]. The input layer is used for preprocessing, standardization, and conversion of data formats, with fixed attributes and no learnable weights. The convolutional layer is the core of the network, which convolves the upper output data and generates new feature maps through activation functions. The convolutional kernel size is used to consider the input data size, receptive field, and network width [18]. The convolutional kernel parameter is a key to learning, affecting the types of features extracted. The relevant calculation is shown in equation (1).

$$x_j^l = f \left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l \right) \quad (1)$$

In equation (1), f represents the activation function. x_j^l represents the j -th output of the l -th layer convolution. M_j represents a subset of feature maps. k_{ij}^l represents the convolutional kernel parameter, and b_j^l represents the bias term. The pooling layer implements downsampling of convolutional feature map. The purpose is to reduce the burden on feature parameters and prevent overfitting to some extent. This process involves multiple pooling strategies, such as average pooling, maximum pooling, and overlapping pooling. The specific operation is displayed in equation (2).

$$x_j^l = f \left(s_j^l \text{down}(x_j^{l-1}) + b_j^l \right) \quad (2)$$

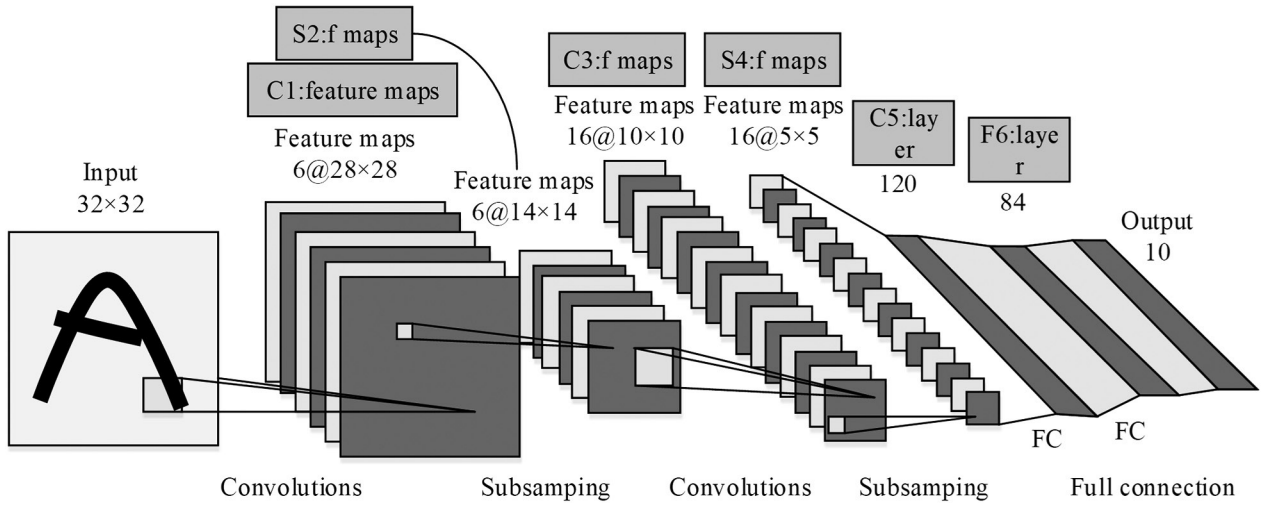


Figure 1. LeNet-5 network architecture.

In equation (2), s_j^l represents the sampling coefficient, and s_j^l represents the pooling process. The fully connected layer is responsible for transforming dispersed feature representations into the sample label space, playing the role of a classifier in the convolutional neural network, as shown in equation (3).

$$x^l = f(w^l x^{l-1} + b^l) \quad (3)$$

In equation (3), x^l represents the output value of the l -th. w^l represents the l -th weight coefficient, and b^l represents the offset term of the l -th fully connected layer.

LeNet-5, originally designed for handwritten digit recognition, has an accuracy rate of 97%, demonstrating excellent classification ability. The goal of this study is to differentiate between non-faulty and faulty characters. This task requires detecting character categories, including uppercase and lowercase letters, numbers, and special symbols.

Handwritten digit recognition typically involves number categories only. However, character defect detection tasks place more emphasis on features related to character size. In light of this, we have optimized the LeNet-5 architecture, which includes adjustments to the network input size (46×32) to maintain the original aspect ratio of characters. A 3×3 small convolutional kernel is used to enhance feature extraction and generalization capabilities. The

ReLU activation function and initialization method are introduced to avoid gradient problems. The improved network structure of LeNet-CB is shown in Figure 2.

Artificial data generation covers operations such as color character image synthesis, random scaling, and rotation of grayscale images, as well as interference factor injection and random breakage. I_{bg} represents the image background. $Chars$ represents the printed character set. The color image synthesis process of label characters is shown in equation (4).

$$G_{dst} = g(\text{rand}(Chars), \text{rand}(I_{bg})), 1 \leq i \leq 100 \quad (4)$$

In equation (4), $\text{rand}()$ represents a random number operation, while $g()$ represents the fusion of characters and background. The random scaling of grayscale images is shown in equation (5).

$$S_{dst} = \text{rand}(\text{gray}(G_{dst})) \times \text{rand}(s), s \in [0.8, 1.2] \quad (5)$$

In equation (5), $\text{gray}()$ represents the grayscale image operation, and s represents the random scaling factor. The grayscale image random rotation and the interference factor injection are shown in equation (6).

$$R_{dst} = \text{rotate}(\text{rand}(S_{dst}), \text{rand}(\theta)) + \text{rand}(\tau), \quad \theta \in [-2, 2], \tau \in [0, 255] \quad (6)$$

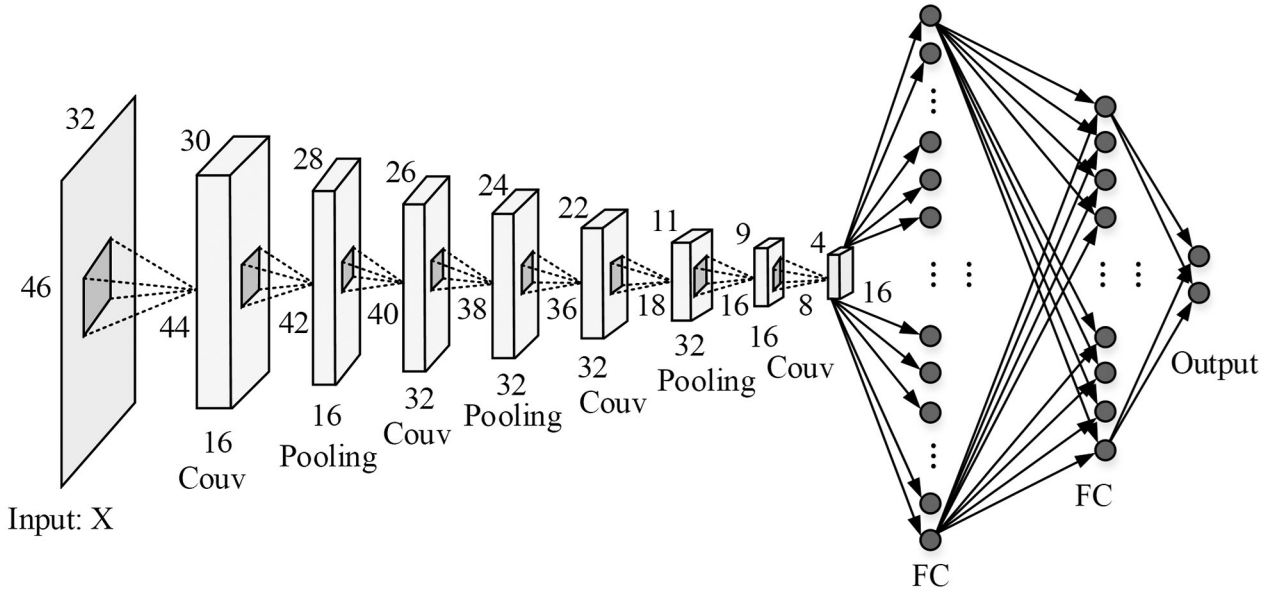


Figure 2. Improved network structure of LeNet-CB.

In equation (6), $rotate()$ represents the rotation function, and θ is the rotation angle. The random damage step is shown in equation (7).

$$B_{dst} = rot(rand(R_{dst}), rand(pos), rand(methods)) \quad (7)$$

In equation (7), $rot()$ represents the damage operation, pos represents the location of the damage point, and $methods$ represents the method set.

3.2. Optimization of LeNet-5 and SENet for Recognizing Electrical Parameters on Labels

While LeNet-5 has demonstrated excellent performance in handwritten digit recognition, its direct application for tasks that involve the recognition of electrical parameters on labels with multiple character categories and high recognition accuracy requirements may not meet all requirements. To address the specific requirements of applying machine vision to intelligent robots for the recognition of electrical parameters on labels, a dedicated electrical parameter label detection and recognition algorithm is optimized based on the LeNet-5 to achieve higher accuracy and stability. In addition, the Squeeze-and-Excitation Network (SENet) enhances the model's representational ability by

introducing feature recalibration mechanisms to strengthen the dependency relationships between feature channels. The SENet model architecture is shown in Figure 3.

In Figure 3, F represents operating on the feature map $X = \{x_1, x_2, \dots, x_C\}$ with C inputs to obtain the feature map $U = \{u_1, u_2, \dots, u_C\}$ of C channels. The convolution operation only slides within the local window of the feature space. Therefore, U does not include dependencies between channels in the previous layer of feature maps. To introduce these dependencies, the SENet introduces the Squeeze and Exception steps. The Squeeze step performs global average pooling on the feature map U in the $H \times W$ space dimension, without changing the same dimension, to obtain $z = \{z_1, z_2, \dots, z_C\}$. The global average pooling is shown in equation (8).

$$z_i = F_{sq}(u_i) = \frac{1}{H \times W} \sum_{m=1}^H \sum_{n=1}^W u_i(m, n) \quad (8)$$

After the operation described by equation (8), the feature map is converted into a channel dependency relationship, as shown in equation (9).

$$s = F_{ex}(z, W) = \sigma(W_2 \cdot ReLU(W_1 \cdot z)) \quad (9)$$

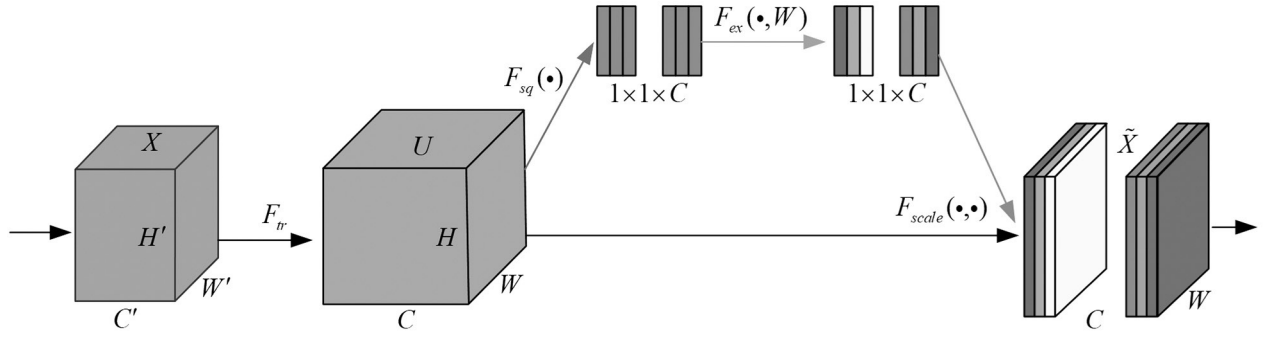


Figure 3. SENet model architecture.

In equation (9), $s = \{s_1, s_2, \dots, s_C\}$, σ represents the sigmoid activation function. The exception step uses the sigmoid function as the activation function, aiming to enable the network to learn the nonlinear relationships between channels while keeping the weight coefficients of each channel independent. Afterwards, the activation value s of each channel is multiplied by the corresponding feature map u for feature adjustment, which assigns weights to each channel in the feature map. Thus, a channel dependent feature map $X = \{x_1, x_2, \dots, x_C\}$ is obtained. The feature adjustment operation is represented by equation (10).

$$x_i = F_{scale}(u_i, s_i) = s_i \cdot u_i \quad (10)$$

The core function of SENet is to enhance the focus on key features and suppress irrelevant features by dynamically adjusting the weight of feature graph channels, to achieve more accurate feature recognition and expression. By introducing this attention mechanism, especially when dealing with complex electrical parameter detection tasks, the recognition efficiency and accuracy have been significantly improved.

Based on the classic LeNet-5 model, this study introduces the Squeeze-and-Excitation Block (SE-Block) [19]. Subsequently, the LeNet with Squeeze-and-Excitation (LENET-SE) model is developed. The LeNet-SE model not only retains the basic architecture of LeNet-5, but also significantly enhances the ability to identify electrical parameter labels through the integration of SE-Block. Based on the improvement strategy of LeNet-CB described in the previous section, the LeNet-SE model has been further optimized. LeNet-CB is a further improve-

ment based on LeNet-5, including adjusting the network input size to 46×32 to maintain the original character aspect ratio, using a small convolution kernel of 3×3 to enhance feature extraction and generalization capabilities, introducing ReLU activation functions and advanced initialization methods to avoid gradient disappearance. Figure 4(a) shows the LeNet-CB structure diagram.

These optimization measures together improve the performance of the LeNet-SE model, making it more accurate and efficient in the detection and identification of electrical parameters. To further improve the performance of the LeNet-SE model, Batch Normalization (BN) is added after each convolution layer to stabilize the learning process and accelerate convergence. The SE blocks are embedded after the third and fourth convolution layers for optimization, as shown in Figure 4(b).

Batch normalization mainly standardizes the output characteristics of convolutional layers. The distribution is adjusted to a standard normal distribution with a mean of 0 and a variance of 1. This processing aims to reduce the tendency of the feature distribution towards the saturation zone of the nonlinear activation function, prevent the gradient vanishing, and thus accelerate the convergence speed of the network. In the batch training stage, implementing BN mainly involves calculating the mean and variance of each batch's features [20]. Then the standardization processing is performed. Taking Figures 3 and 4 as examples, the mean of the feature map can be calculated, as shown in equation (11).

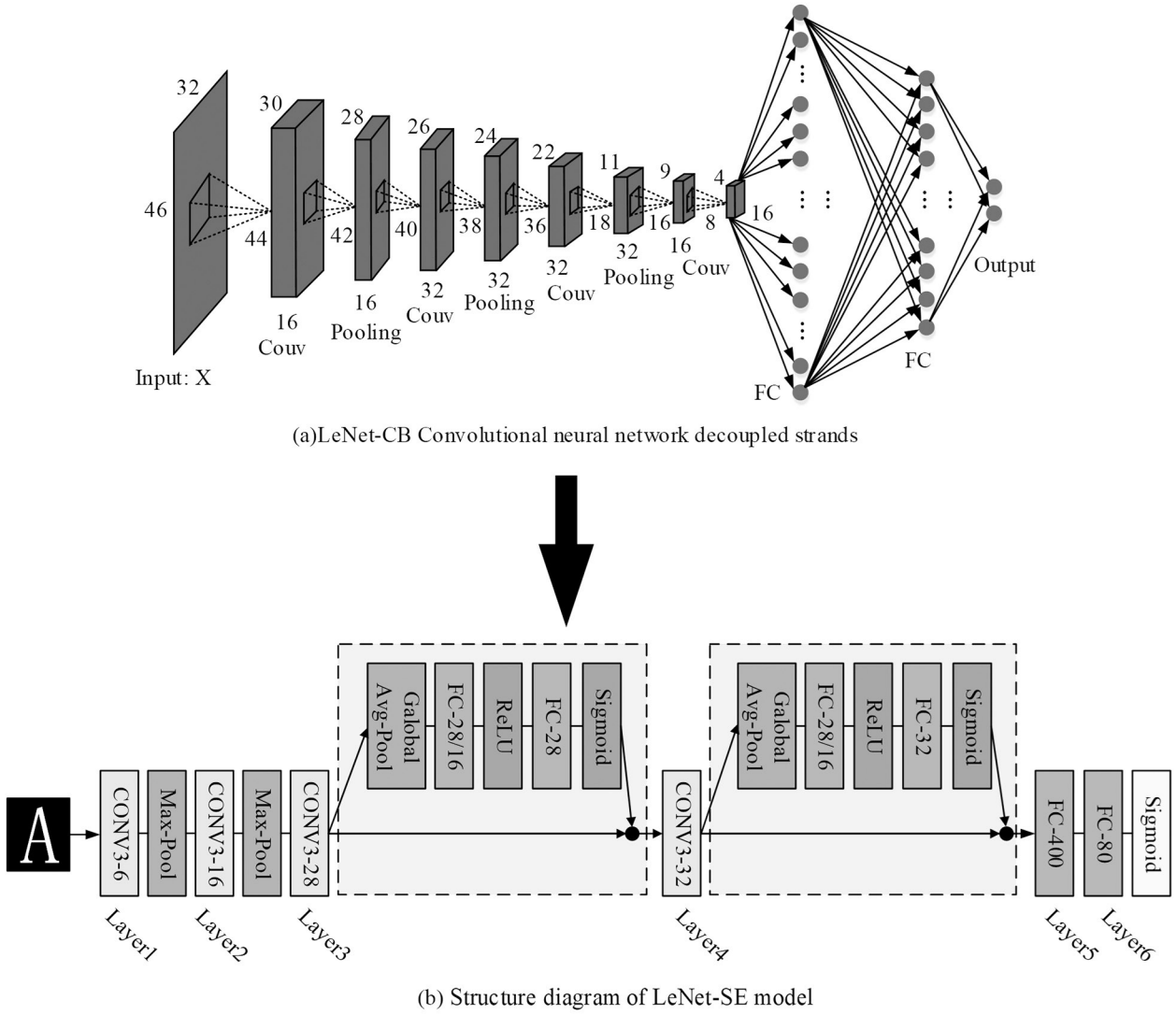


Figure 4. Structure diagram of LeNet-SE model.

$$\bar{X} = \frac{1}{C'} \sum_{i=1}^{c'} x_i \quad (11)$$

For the i -th channel feature map in feature map X , the BN operation is shown in equation (12).

$$\begin{cases} \sigma_x^2 = \frac{1}{C'} \sum_{i=1}^{c'} (x_i - \bar{X})^2 \\ x_i = \frac{x_i - \bar{X}}{\sqrt{\sigma_x^2 + a}} \end{cases} \quad (12)$$

In equation (12), a represents a constant. The BN operation transforms the feature distribution into a standard distribution, which to some extent constrains the expressive ability. To en-

hance this ability, two learnable adjustment parameters γ and β are introduced during the execution of BN, as shown in equation (13). This is crucial for feature expression in robot intelligence.

$$y_i = BN_{\gamma, \beta}(x_i) = \gamma x_i + \beta \quad (13)$$

3.3. LSTM-based Framework for Label Content Parsing and Character Detection Optimization

Due to the lack of accurate identification for electrical parameters in labels, we introduce a fusion of prior knowledge and contextual information based on LeNet-5 to improve the recognition accuracy of easily mixed characters [21].

This is a key element in the machine vision detection algorithms for intelligent robots.

The targeted application of prior knowledge is pivotal in this context. The incorporation of contextual information requires converting strings into sequential data and establishing a neural network capable of processing this data. The LSTM and Gated Recurrent Unit (GRU) are commonly used networks for processing sequential data. Constructing and training of Recurrent Neural Networks (RNN) for this purpose is time-consuming and expensive. Therefore, transfer learning technology is adopted for transfer training and deployment of LSTM-based frameworks, as shown in Figure 5 [22].

In Figure 5, the framework consists of four parts - input, preprocessing, model processing, and output, receiving grayscale or binary images. The preprocessing layer adopts text layout analysis and digital image technology to achieve character segmentation and extraction. The model processing layer loads the trained LSTM network, processes character features, and optimizes recognition results, while the output layer presents the results in text format [23]. LSTM is an upgraded version of the recurrent neural network. The LSTM memory

model is used to manage the transmission and importance of feature information, ensuring the effective transmission of features in a recurrent network. The basic structure is shown in Figure 6.

In Figure 6, there are three control mechanisms and one fast channel within the LSTM unit. The discard structure is responsible for eliminating noncritical information from previous states. The receiving structure determines the contribution of the current input information to the state. The emission structure controls the amount of information from the state to the output [24]. The fast channel quickly transmits the status of each cycle unit to achieve short-term memory function. The calculation method for each control structure is shown in equation (14).

$$\begin{cases} f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\ i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \end{cases} \quad (14)$$

In equation (14), W_f , W_i and W_o represent the weights of the discard structure, receive, and

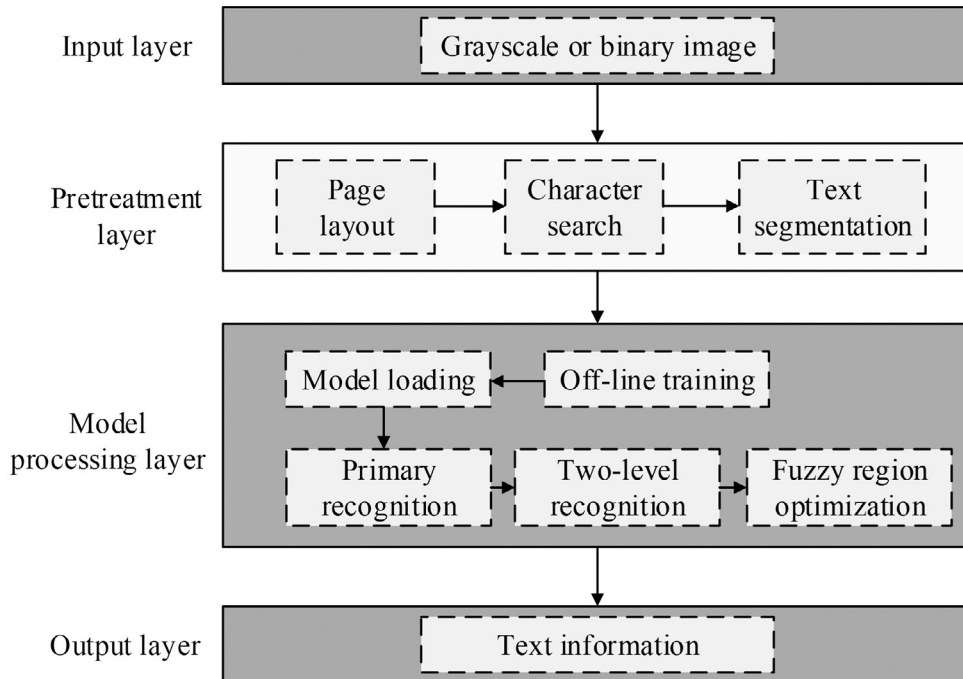


Figure 5. Frame model based on LSTM.

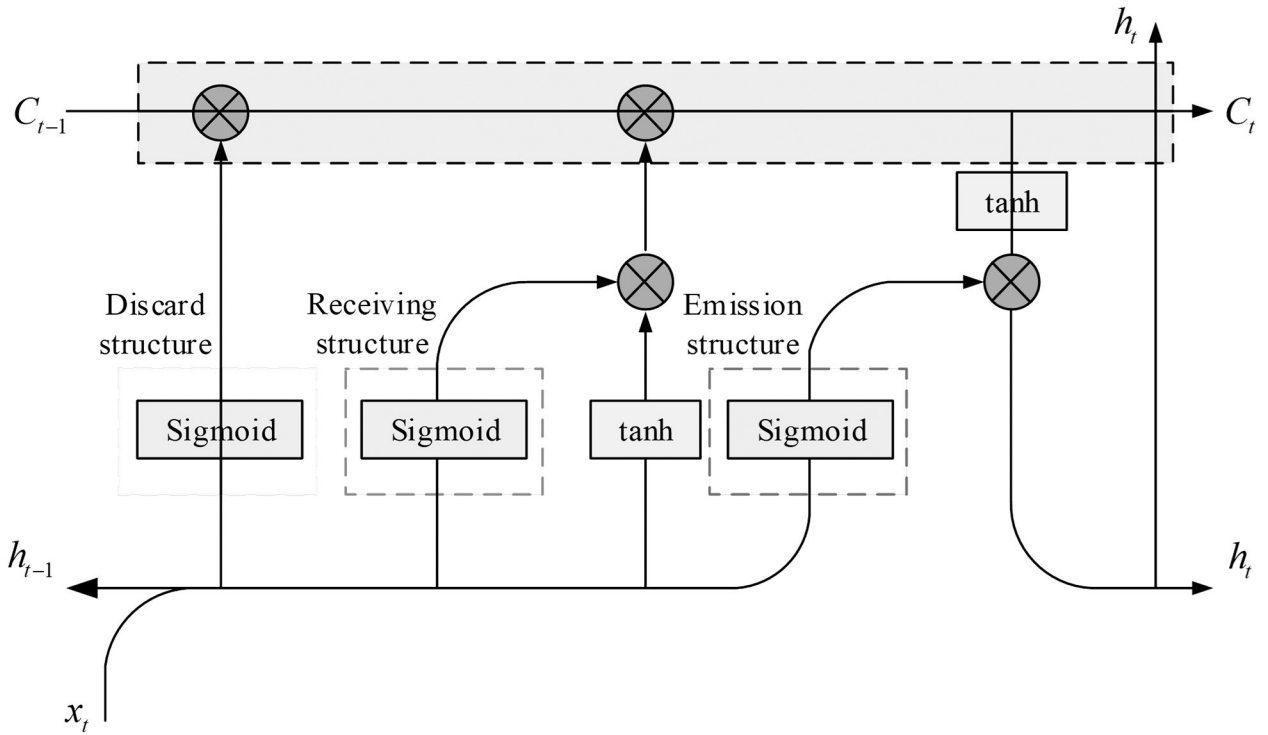


Figure 6. LSTM cell structure diagram.

transmit structure gates. The current unit status and output are shown in equation (15).

$$\begin{aligned} C_t &= f_t * C_{t-1} + i_t \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \\ h_t &= o_t * \tanh(C_t) \end{aligned} \quad (15)$$

In equation (15), W_C and b_C are the weights and biases of the current unit state. C represents current state unit output. h_t represents the output characteristics generated by the current unit. Each LSTM unit generates the current unit state and output after receiving the previous unit state and output as well as the current input [25]. These states are quickly transmitted to the next unit, allowing the network to obtain contextual information of long and short time series, thereby promoting the identification of tag electrical parameters that rely on such information [26]. It has practical value in robot intelligence applications.

4. Result and Discussion

The primary goal of the evaluation is to explore the accuracy of character defects detection method. This involves a comparative analysis of the accuracy between LeNet-5 and LeN-

et-CB. Subsequently, based on LeNet-5 and SENet, the performance of electrical parameter identification and detection is verified. To further improve the accuracy of detection and recognition, the experiment uses standard word libraries, self-trained word libraries, and a combination of the two for electrical parameter detection.

4.1. Results and Analysis of Character Defect Detection Method

A character image dataset is created in the research experiment. 25000 images are developed, named CNN-25K, with a ratio of approximately 1:1 for complete and broken characters. Due to the limited diversity and acquisition cost of industrial data, 1999 complete characters, 799 broken characters, and 99 background images were selected [27]. The data synthesis is conducted using algorithms designed for printing font information. To evaluate the performance of LeNet-5 and LeNet-CB networks in character breakage detection tasks, the two networks are trained on the CNN-25K dataset. The dataset is divided into training, validation, and testing sets in a ratio of 8:1:1. 80% of each type of sample is used for training and the rest

for validation and testing. The specific training results are shown in Figure 7.

In Figure 7, LeNet-5 achieves an accuracy of 98.85% in character detection tasks, which is consistent with the performance in handwritten digit recognition, demonstrating the generalization ability in character classification tasks. The improved LeNet-CB achieves an accuracy of 99.75% in this task, which is 0.9% higher than LeNet-5, confirming the effectiveness of the improved algorithm. Figure 7 also reveals the changes in loss values during the training of the two networks. The loss value after convergence of LeNet-CB is lower than that of LeNet-5. To evaluate the detection performance of the two networks, a faulty character detection test is conducted. The results are shown in Table 1. When the actual measurement accuracy of two networks is lower than that of training, it indicates the possibility of model optimization. LeNet-CB outperforms LeNet-5 in accuracy. Although it takes slightly longer, the difference

is small. Therefore, if hardware conditions are met, priority should be given to selecting LeNet-CB with higher accuracy.

4.2. Performance Analysis of Electrical Parameter Detection and Identification Network Model

The study first selects 12,500 image datasets in CNN-25K. The dataset contains 69 different categories, covering numbers, uppercase letters, lowercase letters, and special symbols. It serves as a standard sample for model recognition. Further, to enhance the generalization ability and prevent overfitting, 51000 images are selected from the publicly available CHARS dataset, involving 62 categories, including various numbers and upper and lower case English letters. These images are fused with the CNN-25K dataset to form a richer LeNet-SE training dataset.

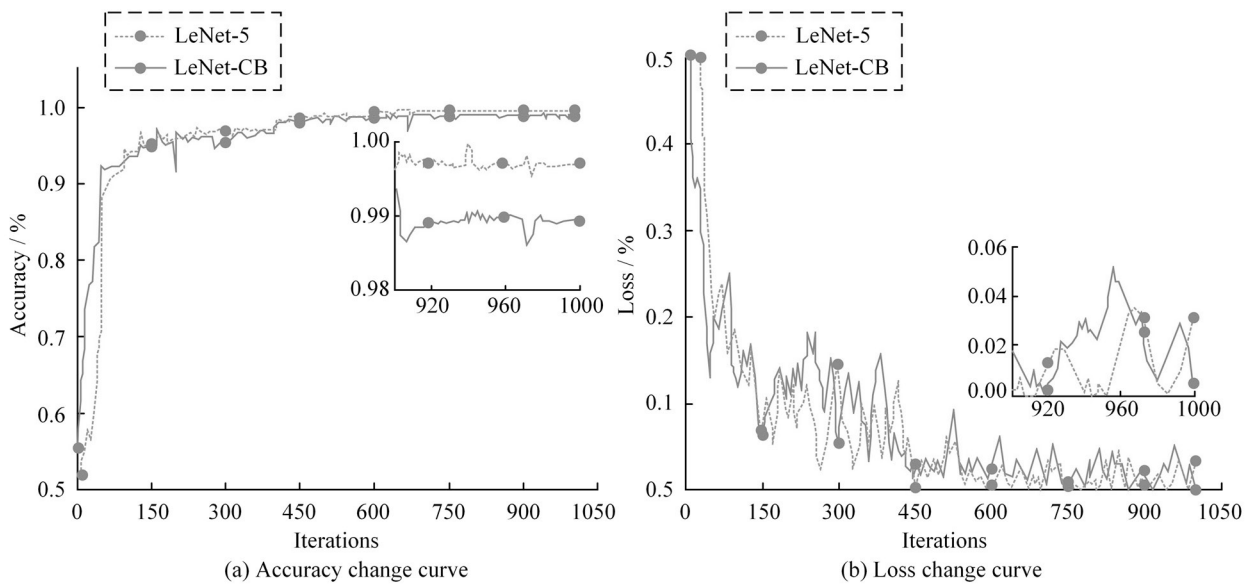


Figure 7. Accuracy and loss value change curves.

Table 1. Comparison of model detection accuracy rate and average duration.

Model	Sample size	Correct number	Accuracy rate (%)	Average duration (ms)
LeNet-CB	999	988	99.49%	10.6ms
LeNet-5	999	979	97.98%	8.57ms

In addition, to ensure that the sample size of various categories in the dataset is as balanced as possible, the data synthesis technology is used to expand the seven special characters. The number of images per special character has increased by about 450. The total amount of data reaches about 65,000 images. The resulting dataset is named CNN-65K, as shown in Figure 8.

To simplify the data loading process of the network model and reduce data annotation and collation work, a concise data organization method is adopted. Each folder in the data set is named using a serial number, representing the corresponding character label. For example, numbers 1-10 correspond to numbers 0-9, numbers 11-36 to uppercase letters A-Z, numbers 37-62 to lowercase letters a-z, and numbers 63-68 to special characters. In this way, the network model can load the path and label of the training data at the same time. During the loading process, the training set, validation set, and test set are divided. The data is randomly shuffled to ensure the effectiveness and reliability of the training process.

To evaluate the performance of network models in electrical parameter identification and detection tasks, LeNet-5, LeNet-CB, and LeNet-SE network models are trained on the CNN-25K dataset. The ablation experiments are conducted on the LeNet-SE network model.

Figure 9 shows the training results of the LeNet-5 and LeNet-CB network models. From Figure 9 (a) it is visible that the training and testing accuracy of the LeNet-5 network model in electrical parameter recognition tasks exceeds 98% after convergence, highlighting the generalization ability of the network model for character recognition tasks. By comparing Figures 9(a) and 9(b), it can be concluded that the training and testing accuracy of the LeNet-CB network model has only slightly improved by about 0.2% compared to the LeNet-5 network model. However, the variance of the former after convergence is smaller. In the neural network model, a small variance means that it has good generalization performance, which verifies the feasibility of the proposed improved algorithm.

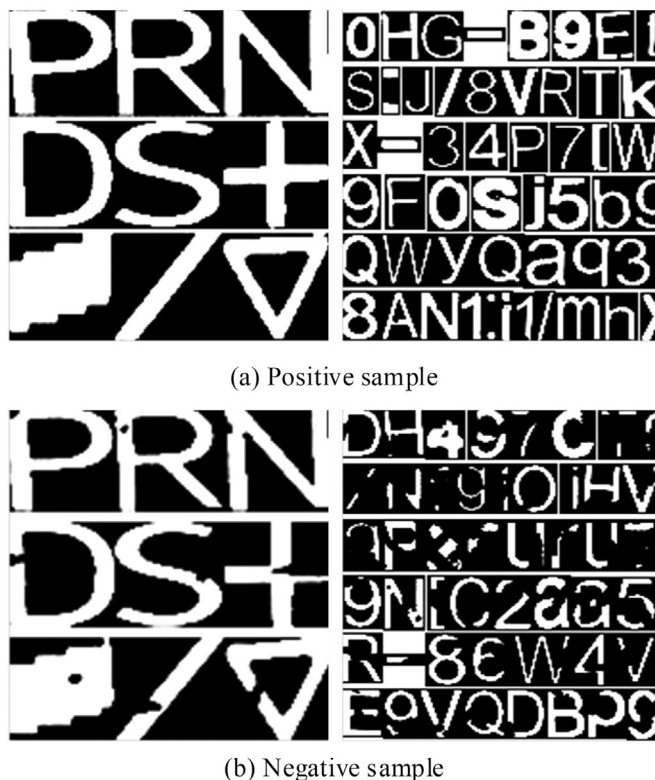


Figure 8. Sample character image.

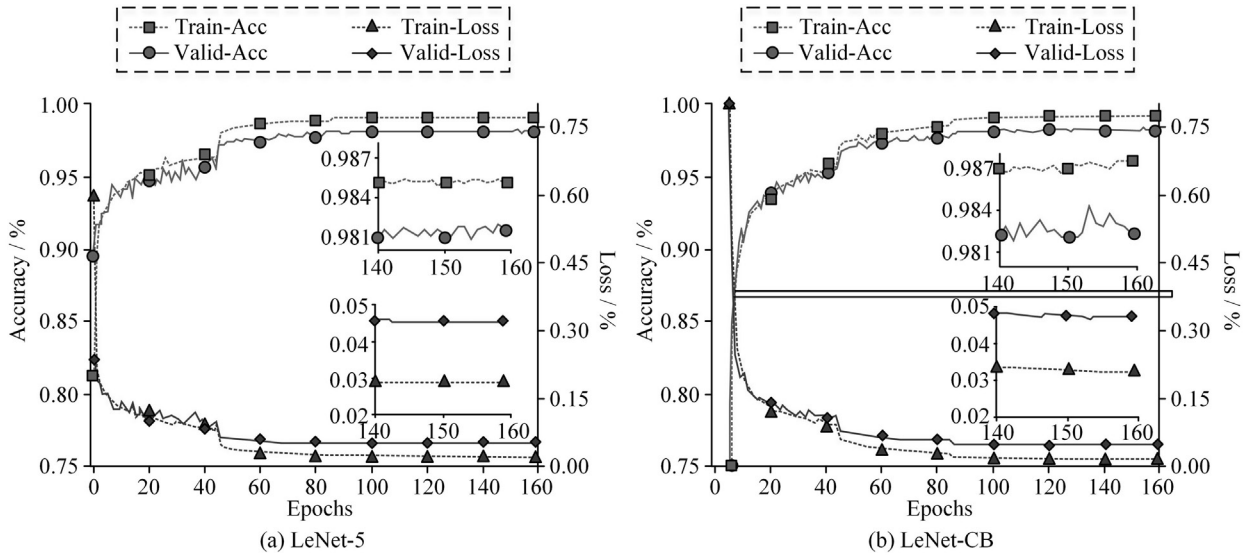


Figure 9. Training results of traditional LeNet-5 and LeNet-CB models.

Figure 10 shows the training results of the LeNet-SE and LeNet-SE without BN network models. By comparing the results between Figures 9 and 10, it is evident that the LeNet-SE surpasses the LeNet-5 network in key indicators such as accuracy, loss value, and variance, verifying the modeling effectiveness of LeNet-SE. Compared with LeNet-CB, LeNet-SE performs similarly in variance. However, it performs better in terms of accuracy and loss value, which indicates that the improved algorithm is practical and effective.

Figure 10 (b) shows the training results after disabling the BN operation. Although disabling the BN operation slightly improves the training accuracy, the testing accuracy and loss performance are not ideal. It proves the important regularization effect of the BN operation. From Figure 9, compared to LeNet-CB, the model with SE Block alone shows an increase of 0.89% in training accuracy and a 3% decrease in loss value, demonstrating the strong potential and positive effects of SE Block.

For further comparison, Figure 11 (a) displays the testing accuracy of LeNet-5, LeNet-CB, and LeNet-SE. LeNet-SE surpasses LeNet-5 and LeNet-CB in testing accuracy, with improvements of 0.5% and 0.98%, respectively, demonstrating the outstanding performance advantage.

To verify the ability of the LeNet-SE network model in electrical character detection, Figure

11 (b) shows the sorted classification accuracy data. Although the LeNet-SE network model has overall high classification accuracy, the detection ability for certain characters is slightly insufficient, as shown in the purple highlighted part in Figure 11 (b). These indistinguishable characters are mainly characters with similar shapes, such as "P" and "p", as well as "X" and "x".

In the labels studied, the same character may have different sizes in different lines. It is impossible to avoid character confusion due to differences in capitalization, size, and similar morphology. To meet the requirements of detection standards, the detection method of electrical parameter labels is further optimized. At the same time, to further evaluate the performance of LeNet-SE network, the commonly used three-layer BP neural network and LeNet-5 network are used as benchmarks for comparative experiments, as shown in Table 2.

The experimental results show that the character recognition speed of LeNet-SE is higher than that of LeNet-5. Compared with BP neural network, LeNet-SE also has a significant improvement in recognition rate, although the difference is still obvious in operation speed due to more network layers. In practical applications, the recognition speed of LeNet-SE is sufficient. The performance can be further improved through subsequent software and hardware optimization.

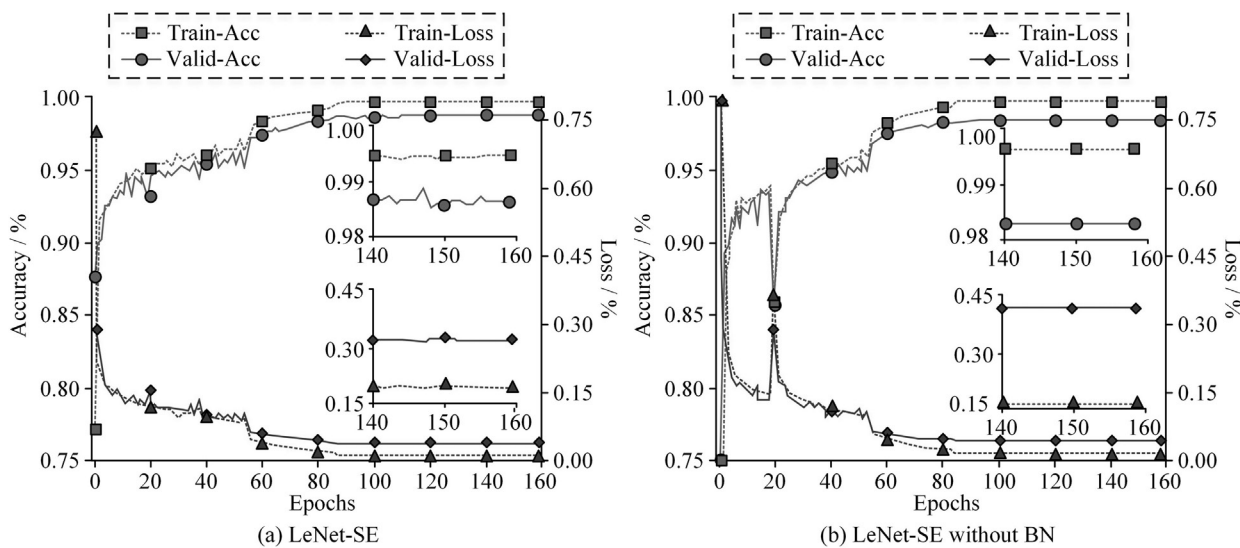


Figure 10. LeNet-SE model training results and experiments.

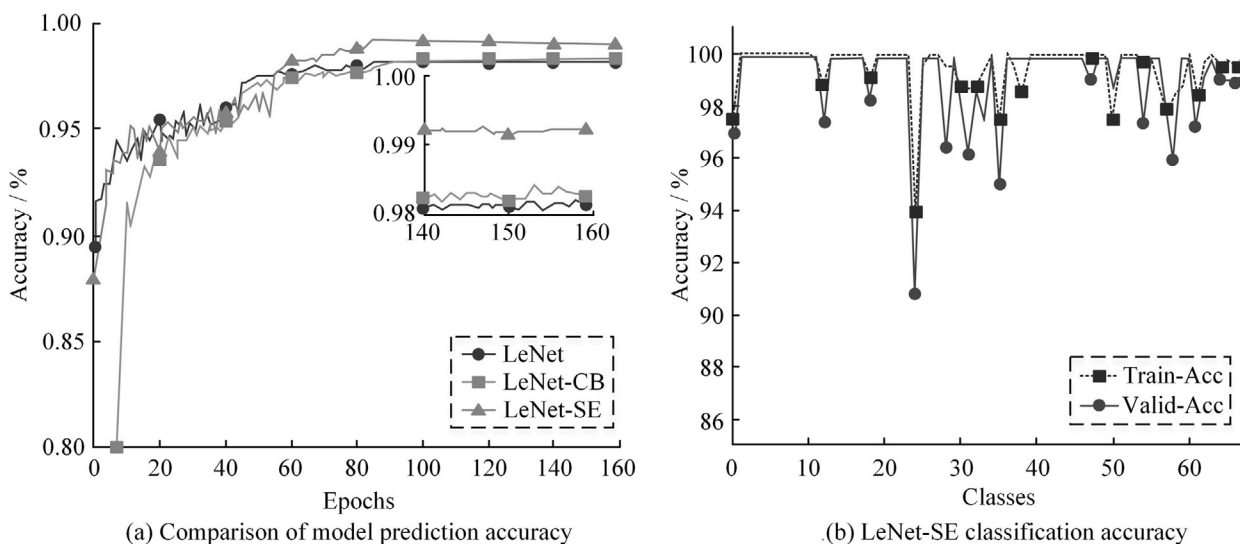


Figure 11. Comparison and prediction accuracy of LeNet-SE classification.

Table 2. Character recognition comparison of the three algorithms.

Network type	Recognition accuracy rate	Average duration
LeNet-5	95%	9.8 ms
LeNet-SE	92%	13.9 ms
BP neural network	88%	4.5 ms

4.3. Detection and Recognition Accuracy, and Running Time Results

To verify the recognition accuracy of algorithm and the effectiveness of transfer learning font files, the experiment uses standard font library, self-training font library, and a combination of the two for electrical parameter detection. In an industrial environment, experiments are conducted on 495 laser engraved label images.

The obtained accuracy data and statistical results are shown in Figure 12. Statistics show that the average detection accuracy using only the standard font library is 91.82%, while the accuracy of the self-trained font library and the combination with the standard font library reach 99.57% and 99.82%, respectively, verifying the effectiveness of the transfer learning font file. From the accuracy curves of the self-trained word library and the combined word library, they show little difference in most electrical parameter detection. It improves detection accuracy and optimizes image quality, reduces image interference, and thus improves the detection and recognition accuracy.

In addition to the detection accuracy rate, the running time of the algorithm is also recorded in this study, as shown in Table 3. The algorithm achieves the best time when using the self-training font file. The average detection time is only 84ms. The result is far below the industry standard time threshold.

In practical applications, such as automated detection systems and real-time monitoring systems, rapid response is crucial. In emergency situations or in high-throughput scenarios, this efficient performance ensures that the system can process large amounts of data in a timely manner, reducing latency and improving overall efficiency. In addition, considering that different application environments may have different computing power configurations, the study further explores the strategy for using standard font and self-training font in high computing power environments. This mixed-use approach aims to optimize the performance of the algorithm in a wider range of application contexts, especially when dealing with complex or diverse data sets.

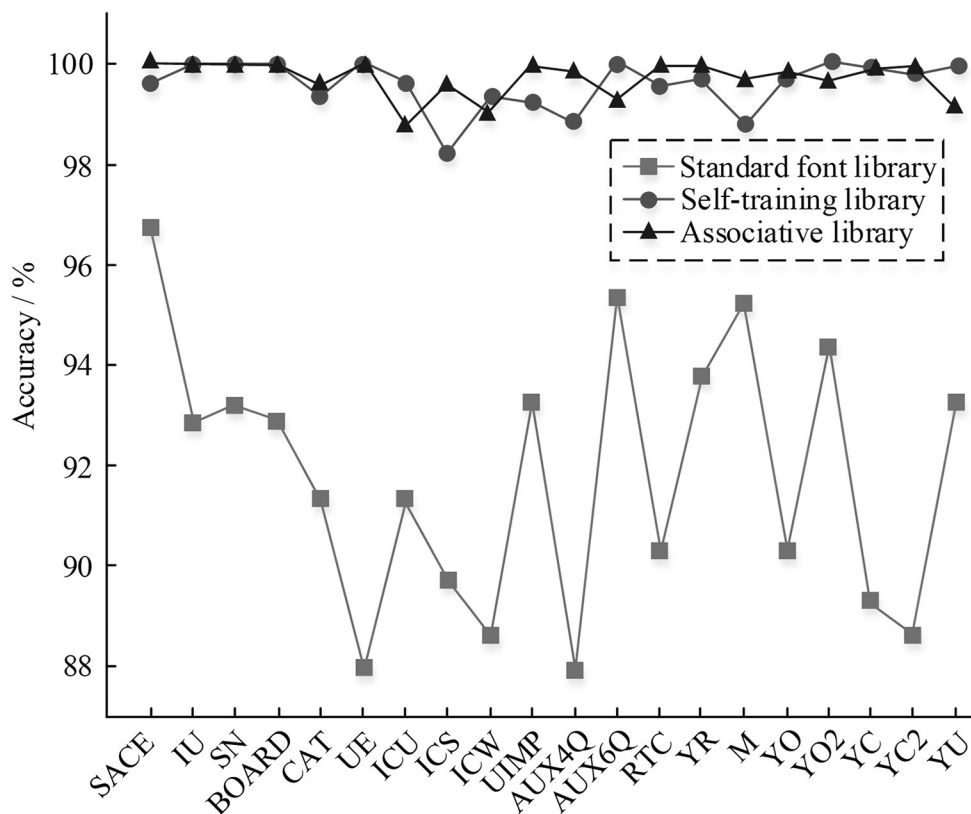


Figure 12. Comparison of tag detection and recognition accuracy.

Table 3. Detection and identification experiment timings.

Self-library	Minimum time consuming (ms)	Maximum time consuming (ms)	Average time consuming (ms)	Average time per parameter (ms)
Standard font library	1628	2759	1893	95
Self-training library	1620	2370	1675	84
Associative library	2119	3130	2317	116

For example, in high-end servers or professional image processing hardware, the combination of the two word libraries can better balance detection speed and accuracy, while improving the ability to recognize various characters and symbols. This flexibility makes the algorithm not only suitable for common configurations, but also scalable to high-end application scenarios that require higher computing power and more accurate detection.

4.4. Test Results of Label Detection and Recognition Algorithm

To comprehensively evaluate the effect and feasibility of the label defect detection method, a comprehensive analysis is conducted on the designed algorithm, as shown in Figure 13.

The next test uses the proposed method to measure the distance between the label wireframe and the backplane baseline. First, the distance between wire frame and baseline is measured manually for the samples with label offset. Then, the proposed algorithm is applied to perform the same measurement. The purpose of this procedure is to compare and verify the accuracy of the algorithm measurements.

To comprehensively evaluate the effect of the algorithm, the results measured by the algorithm are compared in detail with the data measured manually, as shown in Table 4. From

these data, there is a high agreement between the results measured by the algorithm and the values measured manually. The average absolute error of the proposed algorithm is 0.25 mm and 0.29 mm, respectively. These results show that the proposed algorithm meets the expected standards in terms of measurement accuracy and detection accuracy. These findings not only prove the effectiveness of the algorithm, but also emphasize the reliability and accuracy in practical applications.

To ensure that the low-voltage circuit breaker visual inspection system meets the needs of practical applications, several key performance indicators of the system are evaluated comprehensively. These indicators include accuracy, timeliness, reliability, and cost-effectiveness, which together determine the effectiveness of the system.

Based on the pixel accuracy calibration results of the system, the pixel accuracy is 0.0437 mm/pixel. Combined with the pixel resolution standard adopted in the research, the detection accuracy of the system is 0.40mm. This result is significantly higher than the minimum detection accuracy requirement set by the system, which is 0.50 mm. It demonstrates the efficient performance of the system in terms of accuracy.

For testing the detection accuracy of the algorithm, five main detection surfaces of low-voltage circuit breakers are selected as evaluation objects, as shown in Table 5.



Figure 13. Label defect detection examples.

Table 4. Label wireframe distance measurements.

Sample label	Manual measurement / [x,y]	Algorithm measurement / [x,y]	error	Mean absolute error
Sample 1	[7.35, 7.60]	[7.25, 7.90]	-0.11/0.21	0.25/0.31
Sample 2	[7.80, 5.27]	[8.02, 5.15]	0.24/-0.11	
Sample 3	[8.75, 6.45]	[8.85, 6.87]	0.11/0.43	
Sample 4	[8.55, 3.25]	[8.97, 3.85]	0.40/0.62	
Sample 5	[6.85, 4.38]	[6.85, 4.25]	0.01/-0.12	
Sample 6	[7.65, 3.15]	[7.88, 3.90]	0.21/0.75	
Sample 7	[6.05, 6.00]	[6.38, 5.65]	0.33/-0.32	
Sample 8	[7.86, 3.25]	[7.31, 3.47]	-0.55/0.22	
Sample 9	[9.35, 6.70]	[9.12, 7.00]	-0.25/0.21	
Sample 10	[8.99, 6.55]	[8.76, 6.40]	-0.21/-0.14	

Table 5. System detection surface detection accuracy test.

Tag detection surface	Average accuracy	Comprehensive accuracy rate
Label front	98.90%	98.74%
TAB right side	99.35%	
Label back	98.70%	
TAB left side	99.75%	
Tag top	97.01%	

From Table 5 it is evident that the average detection accuracy of the proposed algorithm for the whole system reaches the predetermined design standard. However, in the case of one specific detection item (Tag top), the average accuracy is only 97.01%, which is slightly lower than the expected standard. Further investigation and analysis reveal the potential cause of this problem. The current configuration of the light source is not ideal for the lighting effect of the deep-hole screw when the assembly quality of the deep-hole screw is evaluated. To solve this problem and further improve the detection accuracy, the following work can consider adding a point light source scheme to improve the detection accuracy of the system.

5. Conclusion

To address the diversity, size differences, and irregular shapes of label characters, an innovative LeNet-5 machine vision method based on neural networks is used for detection of label character defects. A label character detection method with a feature map channel mechanism is developed by integrating LeNet-5 and SENet. Additionally, a label content parsing scheme based on LSTM is constructed to enhance classification performance and better serve in intelligent robot applications.

According to the experimental results, the accuracy of LeNet-CB testing reached 99.75%. The average time for single character detection was only 17.9 milliseconds, which meets the detection requirements. For the LeNet-SE detection model with feature graph mechanism, the average recognition and detection accuracy was 98.95%. However, the network had limitations in classifying easily confused characters such as "P" and "p", "X" and "x", which led to the final construction of a label content detection scheme based on the LSTM framework. This method showed that the average accuracy for detection of electrical parameters on labels was 99.57%. The average time for single parameter detection was 84 milliseconds, which meets the detection standards.

Comprehensive testing showed that this method could meet the accuracy and response times requirements. However, there is still room for improvement. Future research will explore detection schemes based on deep learning technology to enhance stability in different operating environments. In addition, the system scheduling logic and robot motion paths will be further optimized to improve system stability and performance, promoting the development of machine vision for intelligent robots.

References

- [1] L. Li *et al.*, "RFID Multi-tag Dynamic Detection Measurement Based on Conditional Generative Adversarial Nets and YOLOv3", *Transactions of the Institute of Measurement and Control*, vol. 43, no. 11, pp. 2472–2482, 2021.
<http://dx.doi.org/10.1177/0142331221991765>
- [2] H. O. Unver and B. Sener, "A Novel Transfer Learning Framework for Chatter Detection Using Convolutional Neural Networks", *Journal of Intelligent Manufacturing*, vol. 34, no. 3, pp. 1105–1124, 2023.
<http://dx.doi.org/10.1007/s10845-021-01839-3>
- [3] R. H. Schmitt *et al.*, "Metrologically Interpretable Feature Extraction for Industrial Machine Vision Using Generative Deep Learning", *CIRP Annals*, vol. 71, no. 1, pp. 433–436, 2022.
<http://dx.doi.org/10.1016/j.cirp.2022.03.016>
- [4] K. K. Sarma and A. Choudhury, "A Vision-Based Framework for Spotting and Segmentation of Gesture-Based Assamese Characters Written in the Air", *Journal of Information Technology Research*, vol. 14, no. 1, pp. 70–91, 2021.
<http://dx.doi.org/10.4018/JITR.2021010105>
- [5] S. Park *et al.*, "Discriminative Feature Learning and Cluster-based Defect Label Reconstruction for Reducing Uncertainty in Wafer Bin Map Labels", *Journal of Intelligent Manufacturing*, vol. 32, no. 1, pp. 251–263, 2021.
<http://dx.doi.org/10.1007/s10845-020-01571-4>
- [6] J. Sethi *et al.*, "NH₂ Linker for Femtomolar Label-free Detection with Reduced Graphene Oxide Screen-printed Electrodes", *Carbon*, vol. 179, pp. 514–522, 2021.
<http://dx.doi.org/10.1016/j.carbon.2021.04.074>
- [7] S. Xiang, "Industrial Automatic Assembly Technology Based on Machine Vision Recognition", *Manufacturing Technology: Engineering Science and Research Journal*, vol. 21, no. 1, pp. 141–148, 2021.
<http://dx.doi.org/10.21062/MFT.2021.018>
- [8] Y. Li *et al.*, "Hybrid Model with Multi-Level Code Representation for Multi-Label Code Smell Detection (077)", *International Journal of Software Engineering and Knowledge Engineering*, vol. 32, no. 11, pp. 1643–1666, 2022.
<http://dx.doi.org/10.1142/S0218194022500723>
- [9] N. Afriliana and A. Ramadhan, "The Trends and Roles of Robotic Process Automation Technology in Digital Transformation: A Literature Review", *Journal of System and Management Sciences*, vol. 12, no. 3, pp. 51–73, 2022.
<http://dx.doi.org/10.33168/JSMS.2022.0303>
- [10] W. Tang *et al.*, "Deep Learning-Based Algorithm for Multi-Type Defects Detection in Solar Cells with Aerial EL Images for Photovoltaic Plants", *Computer Modeling in Engineering and Science*, vol. 1, no. 3, pp. 1423–1439, 2022.
<http://dx.doi.org/10.32604/cmescs.2022.018313>
- [11] C. Zheng *et al.*, "GlassNet: Label Decoupling-based Three-stream Neural Network for Robust Image Glass Detection", *Computer Graphics Forum: Journal of the European Association for Computer Graphics*, vol. 41, no. 1, pp. 377–388, 2022.
<http://dx.doi.org/10.1111/cgf.14441>
- [12] S. E. Chung and H. Y. Ryoo, "Gesture Design Attribute and Level Value of Social Robot: A User Experience Based Study", *Journal of System and Management Sciences*, vol. 10, no. 2, pp. 108–121, 2020.
<http://dx.doi.org/10.33168/JSMS.2020.0208>
- [13] I. Potapenko *et al.*, "Detection of Oedema on Optical Coherence Tomography Images Using Deep Learning Model Trained on Noisy Clinical Data", *Acta Ophthalmologica*, vol. 100, no. 1, pp. 103–110, 2021.
<http://dx.doi.org/10.1111/aos.14895>
- [14] J. A. Kumar *et al.*, "Gender-based Multi-aspect Sentiment Detection Using Multilabel Learning", *Information Sciences: An International Journal*, vol. 606, pp. 453–468, 2022.
<http://dx.doi.org/10.1016/j.ins.2022.05.057>
- [15] Q. Wang and C. Li, "Incident Detection and Classification in Renewable Energy News Using Pre-trained Language Models on Deep Neural Networks", *Journal of Computational Methods in Sciences and Engineering*, vol. 22, no. 1, pp. 57–76, 2022.
<http://dx.doi.org/10.3233/JCM-215594>
- [16] M. Turkoglu, "COVID-19 Detection System Using Chest CT Images and Multiple Kernels-Extreme Learning Machine Based on Deep Neural Network", *IRBM*, vol. 42, no. 4, pp. 207–214, 2021.
<http://dx.doi.org/10.1016/j.irbm.2021.01.004>
- [17] N. Afriliana and A. Ramadhan, "The Trends and Roles of Robotic Process Automation Technology in Digital Transformation: A Literature Review", *Journal of System and Management Sciences*, vol. 12, no. 3, pp. 51–73, 2022.
<http://dx.doi.org/10.33168/JSMS.2022.0303>
- [18] R. Zhou *et al.*, "Learning from Unlabelled Real Seismic Data: Fault Detection Based on Transfer Learning", *Geophysical Prospecting*, vol. 69, no. 6, pp. 1218–1234, 2021.
<http://dx.doi.org/10.1111/1365-2478.13097>

- [19] Z. Chen, "Research on Internet Security Situation Awareness Prediction Technology Based on Improved RBF Neural Network Algorithm", *Journal of Computational and Cognitive Engineering*, vol. 1, no. 3, pp. 103–108, 2022.
<http://dx.doi.org/10.47852/bonviewJCCE149145205514>
- [20] S. Nimrah and S. Saifullah, "Context-Free Word Importance Scores for Attacking Neural Networks", *Journal of Computational and Cognitive Engineering*, vol. 1, no. 4, pp. 187–192, 2022.
<http://dx.doi.org/10.47852/bonviewJCCE2202406>
- [21] C. Zhu, "Intelligent Robot Path Planning and Navigation based on Reinforcement Learning and Adaptive Control", *Journal of Logistics, Informatics and Service Science*, vol. 10, no. 3, pp. 235–248, 2023.
<http://dx.doi.org/10.33168/jliss.2023.0318>
- [22] Y. Guo *et al.*, "Spam Detection Using Bidirectional Transformers and Machine Learning Classifier Algorithms", *Journal of Computational and Cognitive Engineering*, vol. 2, no. 1, pp. 5–9, 2022.
<http://dx.doi.org/10.47852/bonviewJCCE2202192>
- [23] S. H. Yang *et al.*, "A Guideline for Personal Service Robot Interface Design", *Journal of Logistics, Informatics and Service Science*, vol. 7, no. 2, pp. 127–140, 2020.
<http://dx.doi.org/10.33168/JLISS.2020.0209>
- [24] S. Kajkamhaeng and C. Chantrapornchai, "SE-SqueezeNet: SqueezeNet Extension with Squeeze-and-excitation Block", *International Journal of Computational Science and Engineering*, vol. 24, no. 2, pp. 185–199, 2021.
<http://dx.doi.org/10.1504/ijcse.2021.115105>
- [25] L. Enamoto *et al.*, "Multi-label Legal Text Classification with BiLSTM and Attention", *International Journal of Computer Applications in Technology*, vol. 68, no. 4, pp. 369–378, 2022.
<http://dx.doi.org/10.1504/ijcat.2022.125186>
- [26] A. P. Angga *et al.*, "Machine Learning Approaches for Precision Medicine: A Review", *Journal of System and Management Sciences*, vol. 13, no. 4, pp. 416–432, 2023.
<http://dx.doi.org/10.33168/JSMS.2023.0425>
- [27] S. H. Yang *et al.*, "A Guideline for Personal Service Robot Interface Design", *Journal of Logistics, Informatics and Service Science*, vol. 7, no. 2, pp. 127–140, 2020.
<http://dx.doi.org/10.33168/JLISS.2020.0209>

Received: October 2023
Revised: December 2023
Accepted: December 2023

Contact addresses:

Min Zheng
Institute of Aeronautical Engineering
Ezhou Vocational University
Ezhou
Hubei
China
e-mail: 13871802728@163.com

MIN ZHENG is an associate professor at the Institute of Aeronautical Engineering, Ezhou Vocational University. He obtained a master's degree in engineering from Huazhong University of Science and Technology in 2006, majoring in the application of electronic and electrical technology. His research interests are mainly in the fields of applied electronic and electrical technology.
