Power operation ticket review system based on text recognition

Xiao Weidong*[a], Wang Qi[b], Wang Hui[b], Yang Yingjie[b]

[a]State Grid Corporation of China North China Branch Dispatching Control Center, Beijing China 100053; [b]Nanjing NARI Information and Communication Technology Co., Ltd. Nanjing Jiangsu China 211106

* Corresponding author: 18434753989@163.com

ABSTRACT

In order to solve the problem that the handwritten font of the power operation ticket is not easy to recognize, and improve the work efficiency of the dispatch operation ticket. In this paper, a text recognition-based approach to power operation ticket review is proposed. First, a text recognition method based on handwriting feature recognition convolutional neural network (CNN) is proposed, which introduces imaginary strokes and multi-directional features as input, and uses a simple average calculation method to obtain the classification results. Then use the image samples of operation tickets in the actual operation and maintenance to conduct experiments to verify the effectiveness of the method proposed in this paper. Finally, a power operation ticket review system based on text recognition is designed.

Keywords: Action tickets, text Recognition, convolutional neural networks, deep learning

1. INTRODUCTION

The scheduling operation ticket is one of the important components of the scheduling work. Except for special cases such as accident handling, all command operations need to be filled out and signed for the "third review" and signature of the scheduling operation ticket, and finally order execution in logical order [1-3]. With the development of new technologies of mobile devices, mobile phone photography has gradually become the main way to obtain document images for power operation tickets because of its convenience and speed, and is widely used in power operation and maintenance sites [4]. Traditional power operation tickets use optical character recognition (OCR) methods to convert pictures containing optical characters into character format. This method can accurately identify the printed text of the operation ticket with clear imaging. However, the operation ticket of handwritten font is different from the printed font, and the writing habits vary from person to person, which lacks normativeness, which increases the difficulty of text recognition [5,6]. Therefore, it is of great significance to quickly and accurately identify the content of the operation ticket and review it to ensure the safe and stable operation of the power grid.

In recent years, the rapid development of deep learning has provided new solutions to the problem of text recognition. Many scholars have proposed text recognition methods based on deep learning [7,8]. Reference [9] proposed a convolutional neural network handwriting recognition method based on TensorFlow, which can more accurately recognize handwritten fonts. Reference [10] proposed GTN (Graphic Deformation Network), which solved the overfitting problem of CNN model training by translating, scaling, rotating, and stretching text images. However, the algorithmic complexity of the above methods is not high, and the accuracy of handwritten font recognition is low [11].

To sum up, in order to improve the review efficiency of power operation tickets, this paper proposes a power operation ticket review system based on text recognition. First, the archived power operation ticket texts are formed into a canonical training set and a non-canonical training set, and the two are merged to form a data set. Second, the training set is input into the CNN audit module for training, and the audit model is obtained. Finally, input the operation ticket to be reviewed into the CNN recognition model based on handwriting features, and input the recognition result into the review model to obtain the predicted text and review result.

2. CHARACTER RECOGNITION MODEL OF ELECTRIC POWER OPERATION TICKET BASED ON HANDWRITING FEATURE CNN

2.1 Convolutional neural network

Convolutional Neural Convolutional Neural Network is inspired by the working mechanism of the visual cortex of the mammalian brain and is a common deep learning model. CNN models generally include input layers, convolutional layers, pooling layers, and fully connected layers [12, 13]. Among them, the main function of the convolution layer is to convolve the convolution kernel with the local receptive field of the input signal, extract the local receptive field features of the input signal, and use the convolution result of multiple input features as the output feature under the action of the activation function vector.

$$\begin{cases} a_i^{l+1}(j) = f\left\{y_i^{l+1}(j)\right\} \\ y_i^{l+1}(j) = K_i^l * x^l(j) + b_i^l \end{cases} \tag{1}$$

In the formula, $a_i^{l+1}(j)$ and $y_i^{l+1}(j)$ are the final output result and convolution operation result of the jth neuron in the ith feature surface of the lth layer respectively; $x^l(j)$ is the lth The jth local receptive field of layer; K_i^l is the weight vector of the ith convolution kernel of the lth layer; b_i^l is the bias of the ith convolution kernel of the lth layer; $f\{\cdot\}$ is the activation function.

The ReLU function is a piecewise function that avoids the gradient saturation effect and is widely used in deep convolutional networks. Its expression is:

$$\begin{aligned} a_i^{l+1}(j) = f\left\{y_i^{l+1}(j)\right\} &= \max\left\{0, y_i^{l+1}(j)\right\} \\ &= \begin{cases} y_i^{l+1}(j) & y_i^{l+1}(j) \geq 0 \\ 0 & y_i^{l+1}(j) < 0 \end{cases} \end{aligned} \tag{2}$$

The role of the pooling layer is to downsample the feature map to filter out redundant parameters in the data. It has the advantages of feature invariance and preventing the network from overfitting. At present, the commonly used pooling methods are mean pooling and maximum pooling, whose expressions are respectively:

$$p_i^{l+1}(j) = \frac{1}{w}\sum_{i=t-(j-1)w+1}^{jw} q_i^l(t) \tag{3}$$

$$p_i^{l+1}(j) = \max_{(j-1)w+1 \leq t \leq jw}\left\{q_i^l(t)\right\} \tag{4}$$

In the formula, $p_i^{l+1}(t)$ is the pooling result of the neuron in the $l+1$h layer; w is the width of the pooling area; $q_i^l(t)$ is the value of the t-th neuron in the i-th feature plane in the l-th layer.

It can be seen from equations (3) and (4) that mean pooling takes the average value of the receptive field of the feature surface as the output. The maximum pooling outputs the maximum value of the receptive field of each feature surface, extracts important features, and ignores secondary factors. The network structure of this paper adopts the maximum pooling layer.

After the input image is extracted as high-level informative features through the processing of convolutional layers and pooling layers, one or more fully connected layers are connected to classify the informative features. Each neuron in the fully connected layer is connected to each neuron in the previous layer and can receive all the local information of the previous layer. The first fully connected layer develops high-level feature vectors into one-dimensional vectors. The second fully connected layer uses the Softmax regression classifier to solve the classification problem and obtain the final recognition result. Its expression is:

$$J(\alpha) = -\frac{1}{M}\sum_{m=1}^{M}\sum_{d=1}^{D}\left(1\{Y_m = d\}\lg\frac{e^{\alpha_d^T X_m}}{\sum_{h=1}^{D}e^{\alpha_h^T X_m}}\right) \tag{5}$$

In the formula, α is the parameter set of the training model; M is the total number of training set samples; (X_m, Y_m) is the training set samples of the model; D is the number of operating state categories of the synchronous motor; $1\{Y_m=d\}$ is the indicator function, when The function value is 1 when the parentheses are true, and 0 otherwise.

In this paper, the cross entropy loss function is used to calculate the error between the true label and the classification result, and its expression is:

$$Loss = H(p,q) = -\frac{1}{m}\sum_{k=1}^{m}\sum_{j}P_k^j \ln Q_k^j \tag{6}$$

In the formula, m is the size of the input mini-batch; j is the target category; P is the true label of the sample; Q is the classification result output by the model.

2.2 CNN model based on handwriting features

The contents of the power operation ticket include operation items, time, issuer, signature of recipient, etc. There are many handwritten fonts. Aiming at the characteristics of handwritten fonts with complex structure, various types and different writing styles [14], this paper proposes a text recognition method for power operation tickets based on CNN. On the basis of traditional CNN, a variety of handwriting features are introduced to replace the original image input, which breaks the limitation of traditional CNN on the learning of original image spatial features and improves the accuracy of handwritten font recognition.

2.2.1 Imaginary strokes

The imaginary Chinese character uses the degree of directional change to calculate the correlation between different strokes, and extracts the change characteristics between different strokes of the same Chinese character, and then realizes the recognition of handwritten fonts[14]. The formula for calculating the degree of direction change dcd is:

$$dcd = (\frac{\theta}{60}+1)\times min(\frac{mq}{q}\times t, 0.5) \tag{7}$$

Among them, θ is the number of angles ($180 \leq \theta \leq 180$) formed by the connection between different strokes, $mq = 64\sqrt{2}$, q is the stroke length, $t=1/8$.

In the process of writing Chinese characters, it includes the characteristics of starting a pen, dropping a pen, and connecting different strokes. If the connected strokes are shorter and the direction change is larger, it is a strong feature. Strong features can effectively identify the writing features of Chinese characters. Taking the word "作" in the electric power operation ticket as an example, its stroke change characteristics are shown in Figure 1. Among them, the feature pixels are marked by red pentagram. In this paper, by comparing the values of dcd of different pixels, the imaginary stroke matrix is calculated and used as the input of the convolutional neural network model.
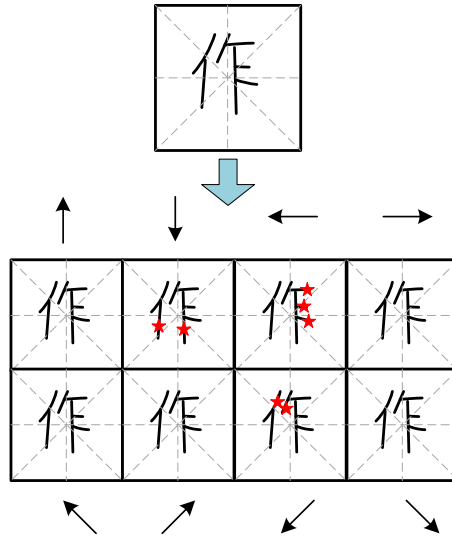
Figure 1.Variation of strokes of the word "作"

2.2.2 8-direction feature

Different from the characters composed of English letters, Chinese characters are mainly composed of horizontal (),
vertical (|), apostrophe (/), and stroking (\), which have obvious directional characteristics [15]. Therefore, multi-
directional features can be used to simulate strokes such as horizontal, vertical, slanting, and stroking of Chinese
characters, and directional features such as 8, 16, and 32 are commonly used. This paper selects 8-direction features, as
shown in Figure 2, respectively from 0°, 45°, 90°, 135°, 180°, 225°, 270°, 315° to calculate the handwriting gradient size,
which can take into account the recognition speed and recognition accuracy. Assuming the start and end coordinates of a
piece of handwriting (x_1, y_1) and (x_2, y_2), the gradient calculation formula is:

$$h_1 = \frac{|d_1 - d_2|}{s} \tag{8}$$

$$h_2 = \frac{\sqrt{2}\min(d_1, d_2)}{s} \tag{9}$$

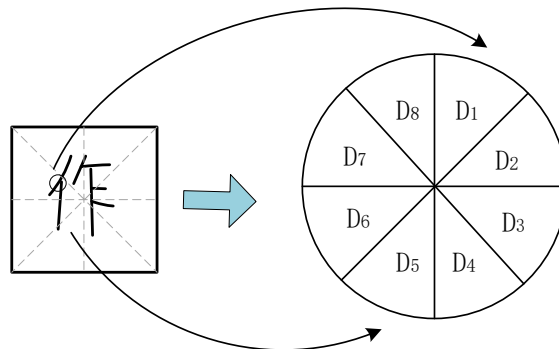Among them, $d_1 = |x_2 - x_1|, d_2 = |y_2 - y_1|, s = \sqrt{d_1^2 + d_2^2}$ 。

Figure 2. Eight-direction characteristics of the word "作"

2.3 CNN model based on handwriting features

The flow chart of the CNN text recognition model based on handwriting features proposed in this paper is shown in Figure 3. The imaginary strokes and 8-direction features of the text to be recognized are respectively used as the feature matrix input model for training, and the simple average method is used to calculate the classification results.

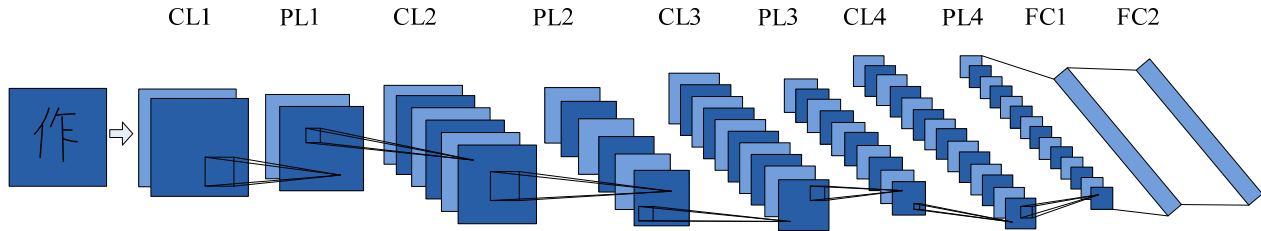CL1   PL1   CL2   PL2   CL3   PL3   CL4   PL4   FC1   FC2

Figure 3. Flow chart of CNN model based on handwriting features

The determination of the size of the convolution kernel has an important impact on the final recognition result. If the size of the convolution kernel is too small, it will increase the network depth and consume a lot of time; if the size of the convolution kernel is too large, the network structure will be too simple, and the features of handwritten text cannot be accurately extracted. , and may contain too much redundant information. The specific structure of the CNN network in this paper is shown in Table 1. The size of the convolution kernel of the first layer is slightly larger than that of the other layers. By extracting short-term features and expanding the receptive field, the useful features of handwritten fonts are effectively extracted and useless information is filtered out; the rest The use of smaller convolution kernels in each layer can reduce network parameters, deepen the number of network layers, enhance computing power and avoid overfitting. In the table, N indicates that the dimensional model of handwriting features is a 6-layer structure, including 4 alternating convolutional layers, maximum pooling layers, 1 fully connected layer and 1 output layer.

Table 1. CNN network structure

Numbering	Network layer	Input size	Convolution kernel size	The number of convolution kernels
CL1	Convolutional layer 1	$90\times90\times N$	$5\times5\times N$	80
PL1	Max pooling layer 1		2×2	
CL2	Convolutional layer 2	$45\times45\times80$	$3\times3\times N$	160
PL2	Max pooling layer 2		2×2	
CL3	Convolutional layer 3	$23\times23\times160$	$3\times3\times N$	240
PL3	Max pooling layer 3		2×2	
CL4	Convolutional layer 4	$12\times12\times240$	$3\times3\times N$	320
PL4	Max pooling layer 4		2×2	
FCL1	Fully connected layer	number of nodes: 512		
FCL2	Output layer	Number of output nodes: 1000		

3. EXPERIMENTAL VERIFICATION

3.1 Experimental results

The experimental verification is carried out with 40,000 images of electric power operation tickets collected by a power supply company of State Grid Power Co., Ltd. as a dataset. Among them, 80% of the images are selected as the training set and 20% as the test set. The dataset includes 1000 commonly used Chinese characters written by 100 writers, each from 100 writers. The experimental environment of this paper is Intel(R) Core(TM) i5-4590, CPU@3.30GHz, RAM24.00GB microcomputer platform, the language used is Python3.7, and the deep learning framework is tensorflow2.7.0. In the training process, the number of iteration steps is 100, and the number of samples in each batch is 100.

In the training phase of the CNN model, the Adam optimization algorithm is used to update the weights, and the learning rate is set to 0.001. The learning rate of the traditional stochastic gradient descent algorithm does not change during the training process, maintaining a single learning rate to update the weights, while the Adam optimization algorithm designs

independent adaptive learning for different parameters by calculating the first-order and second-order moment estimates of the gradient It has strong robustness to hyperparameters and is widely used in the field of deep learning. In order to avoid overfitting the data, the Droupout regularization rule is introduced in the fully connected layer, and the dropout ratio after each convolutional layer is set to 0, 0, 0.05, and 0.1 respectively during the experiment.

Figures 4-5 show the change of loss function and classification accuracy of CNN text recognition based on handwriting features.
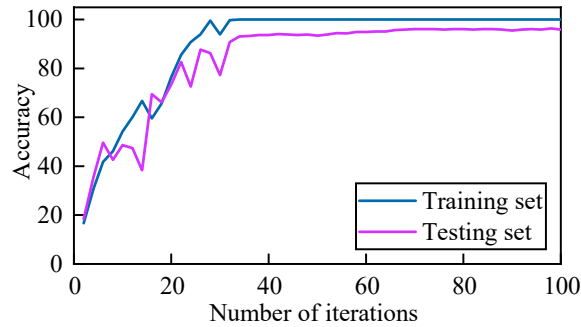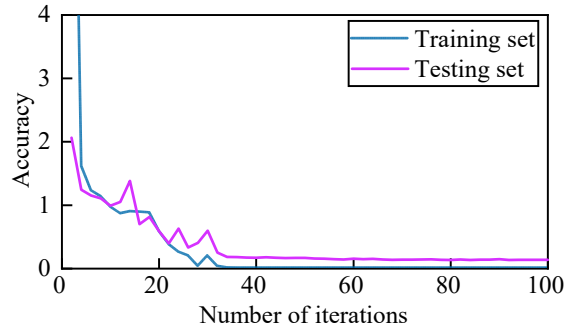

Figure 4. Accuracy change graph

Figure 5. Loss function change graph

As shown in Figure 4-5, as the number of iterations increases, the loss function gradually decreases and converges at 40 iterations, and the accuracy gradually increases, and finally stabilizes at 95.9% at 40 iterations, which can recognize handwriting more accurately Electricity operation ticket.

3.2 Comparison of different algorithms

In order to verify the effectiveness of the method proposed in this paper, a traditional CNN (CNN-none), a CNN model containing only imaginary strokes (denoted as CNN-hy), and a CNN with only directional features (denoted as CNN-di) were selected for comparison. The power operation ticket recognition accuracy of different algorithms is shown in Table 2.

Table 2. Different algorithms text recognition accuracy

Algorithm	Accuracy
CNN-none	80.23%
CNN-hy	87.5%
CNN-di	88.98%
CNN-di+hy	95.9%

As shown in Table 2, the CNN-di+hy method proposed in this paper significantly outperforms other algorithms. Compared with the CNN-none method, the recognition accuracy is increased by 15.67%, and the accuracy is increased by 8.4% and 6.92% compared with CNN-hy and CNN-dy, respectively. It shows that the information obtained based on different handwriting features has strong complementarity, and the fusion of different handwriting features can obtain better classification results.

4. SCHEDULING OPERATION TICKET REVIEW SYSTEM

The power operation ticket review system based on text recognition is mainly composed of three parts: data reading, model training, and text recognition review. The system flow is shown in Figure 6.

(1)Data reading: first collect the archived electric power operation tickets, read the text format for storage, the data set is a normative data set. Then, text processing is performed on the canonical data set, and some texts are generated into irregular data sets through text processing methods such as random interpolation and character deletion. Finally, the normalized data set and the non-standardized data set are merged, and the data is shuffled with random seeds to obtain the data set required for the power operation ticket.

(2)Model training: First, convert the dataset into a code that can be recognized by the computer, and divide the dataset into a test set and a validation set in a ratio of 8:2. Then the dataset is input into the CNN model for training, and after the training is completed, the final audit model is generated and saved.

(3)Text recognition review: collect the images of the power operation tickets to be reviewed, and use the CNN-di+hy model proposed in this paper for text recognition. And save it as a text format backup, input the saved final model, and generate the suggested text. Compare the proposed text with the text to be reviewed, and give suggestions for revision if there is any inconsistency.
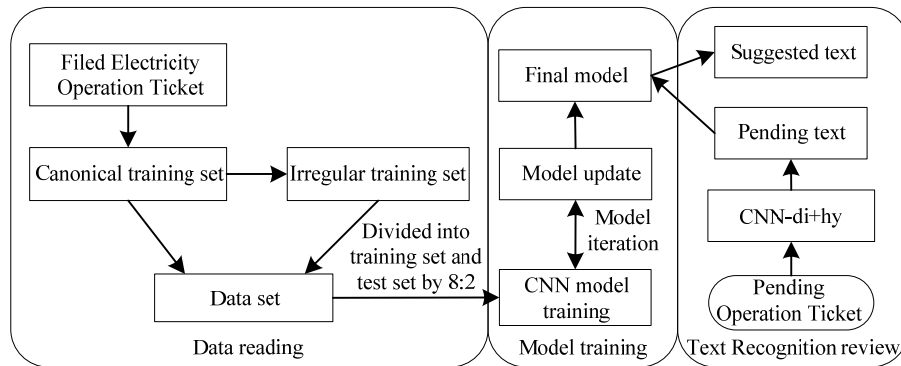

Figure 6. Electricity operation ticket review flow chart

5. CONCLUSION

In this paper, a power operation ticket review system based on text recognition is proposed for the "third review" and signature stage of the power operation ticket. It can realize the accurate identification of the operation ticket image, and improve the work efficiency of the "third trial" of the electric operation ticket. This method has the following characteristics:

(1) Handwriting features such as imaginary strokes and 8-direction features are introduced to obtain richer feature information and improve the accuracy of handwritten text recognition.

(2) The dispatching operation ticket review system based on text recognition proposed in this paper can be used in the "third review" stage of power operation tickets, reducing the review time, improving the accuracy of operation tickets, and ensuring the safe and stable operation of the power grid.

REFERENCES

[1] Zhang Min, Zhao Hailin., "Design and Application of Network Command System for Power Grid Dispatching Operation," Adhesion,2020,44(12):140-144.
[2] Wang Ke, Yao Jianguo, Yu Peiyao, Yang Shengchun, Zhong Haiwang and Yan Jiahao., "Architecture and Key Technologies of Intelligent Decision-making of Power Grid Look-ahead Dispatch Based on Deep Reinforcement Learning," Proceedings of the CSEE,2022,42(15):5430-5439.DOI:10.13334/j.0258-8013.pcsee.220052.

[3] Wu Zibo, Wang Bo, Chen Qing, Guo Yaosong, Zhao Jinhu and Shan Xin., "Research and Application of Dispatch Operation Behavior Mining and Recommendation Technologies Based on Machine Learning," Automation of Electric Power Systems,2022,46(08):181-188.

[4] Liu Yangshao., "Research and application of intelligent operation order system for power grid dispatching," Shandong University,2021.DOI:10.27272/d.cnki.gshdu.2021.006977.

[5] Wang Ning, Zhang Zhimin, Guan Zhichao., "Abnormal data modified model of power grid operation based on OCRtechnology," Information Technology,2022(07):165-170.DOI:10.13274/j.cnki.hdzj.2022.07.030.

[6] Zhang Tingting, Ma Mingdong and Wang Deyu., "Research on OCR Technology,". Computer Technology and Development,2020,30(04):85-88.

[7] Liu Yanju, Yi Xinhai, Li Yange, Zhang Huiyu and Liu Yanzhong., "Application of Scene Text Recognition Technology Based on Deep Learning：A Survey," Computer Engineering and Applications,2022,58(04):52-63.

[8] Gong Faming, Liu Fanghua, Li Juejin and Gong Wenjuan., "Scene Text Detection and Recognition Based on Deep Learning," Computer Systems & Applications,2021,30(08):179-185.DOI:10.15888/j.cnki.csa.008038.

[9] Wan Ruyue, Hai Lin, Gu Zhen., "Handwritten Word Recognition Based on Deep Learning," Modern Information Technology,2021,5(19):89-91+96.DOI:10.19850/j.cnki.2096-4706.2021.19.022.

[10] LECUN Y，BOTTOU L，BENGIO Y，et al., "Gradient-based learning applied to document recognition," Proceedings of the IEEE，1998，86（11）:2278-2324.

[11] CIRESAN D C，MEIER U，GAMBARDELLA L M，et al., "Convolutional neural network committees for handwritten character classification," //2011 International Conference on Document Analysis and Recognition.New York:IEEE，2011.

[12] Xiao Xiong, Wang JianXiang, Zhang Yongjun, et al., "A two-dimensional convolutional neural network optimization method for bearing fault diagnosis," Proceedings of the CSEE, 2019, 39(15):4558-4568. (in Chinese)

[13] Sun Shuguang, Li Qin, Du Taihang, Cui Jingrui, Wang Jingqin., "Fault Diagnosis of Accessories for the Low Voltage Conventional Circuit Breaker Based on One-Dimensional Convolutional Neural Network," Transactions of China Electrotechnical Society, 2020, 35(12):2562-2573.

[14] Zhou Jing, Fang Guisheng., "Judging and fitting method for fractured freehand multi-stroke based on tolerance zone," Computer Science,2017,44(S2):184-188.

[15] BAI Z L，HUO Q., "A study on the use of 8-directional features for online handwritten Chinese character recognition," //Eighth International Conference on Document Analysis and Recognition.New York:IEEE，2005.