Research on Image Recognition of Electrical Equipment based on Deconvolution Feature Extraction

Zhe Li^{1*}, Haifeng Su²

¹North China Electric Power University, Department of Electrical Engineering, Lianchi, Baoding, Hebei, China

²North China Electric Power University, Department of Electrical Engineering, Lianchi, Baoding, Hebei, China

Abstract. Based on machine learning technology and combining the operation of machine learning from the idea of neural network, this paper focuses on the classification and recognition of image data of transformers, circuit breakers and isolation switches in substations. Firstly, the image enhancement is carried out on the basis of the original image, which simulates the possible scenes in reality. Secondly, using the dual-mode a deconvolutional network to capture significant features from in-depth visible and infrared images. Furthermore, all these features are subjected to the program to conduct transfer learning and weighted fusion. The dual-mode deconvolutional network (DMDN) extracts and highlights the features of the electrical equipment. Compared to traditional model, the recognition accuracy of the improved model is reached at 99.17%.

1 Introduction

With the rapid development of camera system and Computer vision technology, power system has gradually entered the intelligent monitoring period [1].

At present, many experts and scholars have achieved satisfactory results from conducting in-depth research in the field of image analysis and recognition of electrical equipment, and have achieved satisfactory results. Reference [2] used deep learning model to extract features of power equipment, and then combines with Random Forest classifier to achieve the image recognition of important electrical equipment. Reference [3] constructed a deep convolutional autoencoder network model based on convolutional neural network and combined with autoencoder. The model automatically learns and extracts effective features in small sample images to improve recognition accuracy. Reference [4] proposed an image recognition method for electrical equipment based on the GoogLeNet Inception-V3 model. This method has a high average recognition accuracy for circuit-breakers, current transformers, insulators, arresters and potential transformer. Reference [5] set up the overall objective function for the deconvolution layer and modifies it to a supervised state, and finds the most suitable high-level feature mapping matrix for the training of the back-end convolutional neural network.

In this paper, for small sample data, the data is enhanced, and then the dual-mode deconvolution network (DMDN) is used to extract the features of the visible light and infrared images of the transformer, circuit breaker, and isolating switch, which are pre-set based on the feature extraction of the traditional convolutional neural network. The aim is improving the

recognition accuracy using the dual-mode network for overall characteristics of the specific small samples and perform transfer learning and feature fusion on the extracted feature matrix.

2 Convolutional neural network learning based on dual-mode deconvolution feature extraction

2.1 Network structure

The structure of the convolutional neural network based on dual-mode deconvolution network (DMDN) feature extraction is shown in Figure 1. The network consists of two parts. The first part is the dual-mode deconvolution feature extraction network. Given the visible light and infrared images of the electrical equipment and the randomly initialized feature mapping matrix, the deconvolution network is used for directly performing the decomposition mapping operation from the original image to the hidden layer feature space. At the same time, the decoder reversely maps the hidden layer features back to the input space to reconstruct the input similar to the original image, and differences between the reconstructed image and the original image are used as the optimization target of the objective function. The second part is the recognition training of convolutional neural network. The dual-mode feature mapping matrix learned in the first stage is used as the initial convolution kernel matrix of CNN, and the extracted initial convolution kernel matrix is weighted and fused to realize cross-channel information interaction and integrate, then perform alternate convolution, activation,

* Corresponding author: m15561436123@163.com

normalization, and pooling operations (CANP). Finally, the Mini-batch Gradient Descent (MB-SGD) is used to fine-tune CNN to avoid the vanishing gradient problem during training.

Figure 1. System structure

2.2 Image feature extraction of electrical equipment

In this paper, the deconvolution layer is used to stack the hierarchical network model, and the efficient optimization technology is used to extract the feature mapping matrix. The reasoning operation principle of deconvolution network model is shown in Figure 2.

Figure 2. Deconvolution feature extraction

Given an input image and a randomly initialized convolutional neural network, the feature extraction operation is to perform deconvolution on the top feature map z_i to obtain the reconstructed image s_i .

Assuming that the image y in the original samples is included in the convolution operation to obtain the feature map, the reconstructed image s_i is obtained by using the deconvolution kernel to do the convolution operation:

$$s_i = \sum_{l=1}^{N_l} z_i^l * f_i^T \quad (1)$$

The ultimate goal of feature extraction is to make the reconstructed image and the input image as similar as possible, so the minimum reconstruction error is defined as:

$$\min C_1(y) = \sum_{i=1}^N \|s_i - y_i\| + \sum_{i=1}^N |z_i| \quad (2)$$

The feature learning process of the deconvolution network can be divided into the following two stages:

the first stage, given the convolution kernel f_i , find the feature map z_i to minimize the objective function. The second stage, given the feature map z_i , solve the convolution kernel f_i , which minimizes the objective function.

In the specific training process, it is necessary to introduce an auxiliary objective variable α_i for the first time to avoid falling into the local optimum, and then solve the minimum value of the auxiliary objective function $C_2(y)$ to make the auxiliary variable and the feature map approximate. The auxiliary objective function is defined as:

$$C_2(y) = \sum_{i=1}^N \|s_i - y_i\|_2^2 + \sum_{i=1}^{N_l} \|z_i^l - \alpha_i^l\|_2^2 + \sum_{i=1}^N |\alpha_i^l| \quad (3)$$

After introducing this auxiliary function, alternately fix the values of z_i and α_i to obtain the optimal solutions respectively. First, fix z_i and solve α_i . Derivation of equation (4):

$$\frac{\partial C_2(y)}{\partial z_i^l} = \sum_{k=1}^{N_l} (F_k^l)^T \left(\sum_{k=1}^{N_l} F_k^l z_i^l - z_i^{l-1} \right) + \mu (z_i^l - \alpha_i^l) \quad (4)$$

In the stage of finding auxiliary variables α_i , if the feature graph z_i is fixed, the optimal problem α_i can be solved by the feature maps. the solution expression α_i of auxiliary variable is:

$$\alpha_i^l = \max \left(\left| z_i^l - \frac{1}{\mu} \right|, 0 \right) \frac{z_i^l}{|z_i^l|} \quad (5)$$

The update process of the feature map z_i is: firstly, given i , set equation (4) equal to zero, and solve z_i , that is, to solve the following dimensional linear system:

$$A \begin{bmatrix} z_1^l \\ \vdots \\ z_i^l \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N_l} (F_{1,c}^l)^T z_1^l + \alpha_1^l \\ \vdots \\ \sum_{i=1}^{N_l} (F_{i,c}^l)^T z_i^l + \alpha_i^l \end{bmatrix} \quad (6)$$

Where,

$$A = \begin{bmatrix} \sum_{i=1}^{N_l} (F_{1,c}^l)^T F_{1,c}^l + \alpha_1^l & \cdots & \sum_{i=1}^{N_l} F_{1,c}^l (F_{i,c}^l)^T \\ \vdots & \vdots & \vdots \\ \sum_{i=1}^{N_l} (F_{i,c}^l)^T F_{1,c}^l & \cdots & \sum_{i=1}^{N_l} (F_{i,c}^l)^T F_{1,c}^l + \alpha_i^l \end{bmatrix}$$

Finally, using the optimal solution of the conjugate gradient descent equation (5) and the convolution kernel can be updated according to the gradient descent method:

$$\frac{\partial C_2(y)}{\partial f_{i,c}^l} = \sum_{i=1}^{N_l} (z_i^l)^T \left(\sum_{k=1}^{N_l} F_k^l f_{i,c}^l - z_i^{l-1} \right) \quad (7)$$

2.3 Convolutional neural network image recognition training

CNN includes convolution layer, pooling layer and full connection layer. Softmax classifier is used for multi task classification. According to the current mainstream CNN training techniques, nonlinear Relu activation function is used to enhance the sparsity and linear separability of features. The calculation of the characteristic images is as follows:

$$z_{j'}^l = f \left(\sum_{j \in M^l} z_j^{l-1} * f_{jj'}^l \right) \quad (8)$$

Where, l is the number of layers of the convolution layer, z_j is the output characteristic graph j of the layer l , f_{jj} is the convolution kernel connecting the first characteristic graph j of the layer $l-1$ and the second characteristic graph j of the layer l , M^l is the number of characteristic graphs of the layer $l-1$, $*$ is the convolution operator.

Because the distribution of input data changes after convolution operation, the standardized BN operation is introduced to process the data distribution. The BN is placed after the convolution operation of the network activation function, and the forward propagation convolution formula is transformed into:

$$z_{j'}^l = f \left(BN \left(\sum_{j \in M^l} z_j^{l-1} * f_{jj'}^l \right) \right) \quad (9)$$

The pooling layer downsamples the feature map of the previous convolution layer to obtain the output feature map with smaller dimension corresponding to the input feature map one by one :

$$z_{j'}^l = f \left(BN \left(\beta_j^l \cdot down(z_j^{l-1}) \right) \right) \quad (10)$$

Where: $down()$ is the down-sampling function and β is the down sampling coefficient.

The CNN model is trained by back propagation layer by layer, and the error cost function is:

$$J(f) = \frac{1}{2} \sum_{i=1}^N \sum_{m=1}^M (\hat{y}_m^i - y_m^i) \quad (11)$$

Finally, MB-SGD is used to update the weights:

$$f_{jj'}^l(t) = f_{jj'}^l(t-1) - \frac{1}{N} \eta \sum_{i=1}^N \frac{\partial J}{\partial f_{jj'}^l(t-1)} \quad (12)$$

Where, t represents the current time and η represents the learning rate.

2.4 Algorithm steps

The input of the model in this paper is a training sample set of N samples, and the output is the classification label and accuracy of the test sample in the fixed sample. The specific experiment process is as follows:

2.4.1 Use deconvolutional network to extract feature mapping matrix

First, use equation (1) to deconvolve the feature map z_i of the input image to obtain the reconstructed map s_i . Then introduce auxiliary variable α_i , use formula (5) to obtain auxiliary variable α_i . Next, according to the obtained auxiliary variable α_i , use equation (6) to obtain the characteristic map, after multiple iterations of the above steps, use equation (7) to update the convolution kernel, and finally output the feature mapping matrix.

2.4.2 Use DMBDCNN for image fusion classification training

At first, the feature map matrix extracted by the deconvolution network is weighted and fused as the initial matrix of the DMBDCNN convolution kernel, and the image data is convolved using formula (8). Secondly, use equation (9-10) to down-sampling and normalize the obtained convolution feature map. After training on each layer of the model, use formula (11-12) to supervise and fine-tune the convolutional network to update the model parameters.

3 Experimental results and analysis

Figure 3 shows the results of some original data after image enhancement. After data enhancement processing, the existing training samples will generate more training data.

Figure 3. Some of the data enhances the presentation

This paper studies the classification and identification of transformer, circuit breaker and disconnector. A total of 1200 images are used for training and testing. The image size is set to 224*224*3 after preprocessing, and each type of equipment has 400 images. During the training, the number of randomly selected images of transformer, circuit breaker and disconnector is 360, and the remaining 40 are taken as test samples.

3.1 Single-mode deconvolution network (SMDN) visible light image recognition

For the classification of visible light electrical equipment images, this paper introduces the deconvolution feature extraction model. Compared with the traditional neural network classification algorithm, the model adjusts the number of convolution kernel layers to match the deconvolution feature extraction layer.

The freezing training method is adopted. Firstly, only the whole connection layer is trained. Secondly, unfreeze

all layers for training. When using small batch to accelerate training, the batch size is set to 10. The Dropout layer with the discard rate of 0.3 is introduced in the full connection layer 1 and 2, so as to avoid over fitting and freeze the learning rate of the first 100 freezing trainings is 0.001, and the learning rate of the last 100 unfreezing trainings is 0.0001. The classification accuracy of visible images is shown in Figure 4.

Figure 4. Average accuracy change graph under different training times

It can be seen from Figure 4 that compared with the traditional AlexNet and VGGNet, the average recognition accuracy of SMDNCNN has reached 97.32%, indicating that the model in this paper can better adapt to the characteristics of fewer electrical equipment samples.

3.2 Dual-mode visible and infrared image fusion recognition

For electrical equipment visible and infrared image fusion classification, this paper introduces a weight coefficient μ after deconvolution feature extraction, uses the weighted fusion method for feature fusion. Different weight coefficients will affect the feature parameter matrix and affect the accuracy of image classification. Table 1 shows the influence of different weight coefficients μ on the classification accuracy rate.

Table 1. Classification accuracy (μ)

μ	0.1	0.5	0.9	1	3	5
Accuracy/%	97.8	98.8	99.17	88.1	33.3	33.3

With the increase of the weight coefficient μ , the classification accuracy gradually improves, and the weight coefficient $\mu = 0.9$ is the critical value. Beyond this value, the accuracy decreases rapidly.

Figure 5. Comparison of average accuracy between SMDNCNN and DMDNCNN

It can be seen from Figure 5 that after 100 freezing training cycles and 100 thawing training cycles, compared with 97.3% of the single mode visible light image recognition accuracy, the accuracy of the dual mode fusion recognition reaches 99.17%. The experimental results show that fusing the features extracted by deconvolution can highlight the unique information of different electrical equipment, and the parameter learning of neural networks can develop in a better direction. The further improvement of the average recognition accuracy rate illustrates this point.

4 Conclusion

Aiming at the local optimization and vanishing gradient problems that may be caused by the random initialization of the convolution kernel and the gradient descent learning method in the CNN training process, this paper proposes a convolutional network learning model based on deconvolutional feature extraction. The dual-mode deconvolution network (DMDN) used in this paper is able to automatically learn the feature mapping matrix of the image, which is used as the initial value of the convolution kernel to participate in the training of the CNN model. In addition, after weighted fusion of the learned visible light and infrared features, the stronger characteristic expression effectively solves the limitation of the limited number of samples. Compared to traditional model, the recognition accuracy of the improved model is reached at 99.17% after analysis and verification.

Acknowledgments

This paper was supported by the National Key R & D Program of China (2017BJ0080).

References

1. Y.Q. Song, G.L. Zhou, Y.L. Zhu. Present Status and Challenges of Big Data Processing in Smart Grid. J. Power System Technology, 37(4): 927-935(2013)
2. J.F. Li, Q.R. Wang, M. Li. Electric Equipment Image Recognition Based on Deep Learning and Random Forest. J. High Voltage Engineering, 43(11): 3705-3711(2017)
3. H.R. Sun, Z.J. Pan, Y. Yan. Identification and location of photovoltaic hot spots with small samples based on deep convolutional autoencoder network. J. Journal of North China Electric Power University, 1-8(2021)
4. K. Xu, Z.J. Ling, Y.Y. Zhang. Image Recognition of Electric Equipment Based on GoogLeNet Inception-V3 Model. J. High Voltage Apparatus, 56(09): 129-135+143(2020)
5. E.H. Lv, X.S. Wang, Y.H. Cheng. Deep convolution neural network learning based on deconvolution feature extraction. J. Control and Decision, 33(03): 447-454(2018)