Accurate Positioning of the Substation Instruments in Images by Using a Method Based on Convolutional Neural Networks

Caiyi Yang, Kun Huang, Yujie Li, Yue Tang and Jianying Yuan*

College of Control Engineering, Chengdu University of Information Technology, Chengdu 610225, China.
*corresponding author, Email: 897455865@qq.com

Abstract. Aiming at solving the problems of sensitivity to instrument shape and low accuracy of positioning by using the traditional feature matching points based methods, a novel positioning approach based on convolutional neural network is proposed in this work. The designed convolutional neural network consists of two convolutional layers, two pooling layers and two fully connected layers. The objective function adopts cross-entropy and the optimization method adopts Adam algorithm. We used 7000 images collected at different times and situations as test samples. We discussed the performance of the proposed algorithm under different amount of training data, different convolution kernel sizes and different number of convolution kernels in the experiments. Compared with the traditional feature matching point based method, the proposed method has higher recognition accuracy and lower false positioning rate.

1. Introduction
In recent years, the construction of smart grid has promoted the development process of unmanned substation, and the substation inspection robot has gradually become a research hotspot. The substation inspection robot reaches the designated position through the autonomous navigation system, detects the equipment status through the sensors. It can replace the traditional manual inspection and feedback the equipment operation status to the station in time. The accurate positioning of the instrument in the images is the premise of reading. Most substation equipment is outside, and the images collected by the robot are easily affected by the light, especially when there is a high-reflective object in the scene, the target image often shows light pollution (such as too bright, too dark, local strong exposure). In addition, due to the inevitable errors in robot autonomous navigation, the positions and angles of images collected at different times are different, which may lead to the untargeted, distorted and fuzzy images. Therefore, it is significant to build a robust instrument positioning algorithm that can resist light changes and image deformation. In this paper, the instrument positioning method of substation based on CNN is proposed, and the influence of different training samples, convolution kernel size and convolution kernel number on the accuracy of instrument positioning are analyzed. The results also showed that the proposed algorithm had higher accuracy than the traditional SURF matching algorithm.

2. Relative Work
In recent years, with the mature development of image recognition technology and the need for intelligent construction of power grids, image recognition based on substation instrumentation technology has become a key research hotspot. Relevant scholars have done a lot of work in this field. A certain result has been achieved. In general, existing research methods can be divided into two
categories. The first category is a meter positioning method based on feature matching. The second is a neural network based on instrument identification method.

In the first type of method, Zhu Bolin et al.[1] collected template images of different instruments. For the instrument images collected, the corresponding template maps are invoked in the template library. ORB algorithm based on the directional binary simple descriptor [2] is used to calculate the homograph matrix of the template and the image to be tested, and then the instrument sub-region is extracted. Fang Hua [3] proposed a method called SIFT algorithm for locating the instrument. Firstly, the original image and the Gaussian convolution kernel are convoluted to generate the scale space, and then SIFT feature vectors were generated by removing scale transformation, rotation and other geometric transformation factors, localization was completed after normalization and template image matching. Under different lighting conditions, the SIFT feature registration algorithm can accurately acquire the instruments in the image acquired. Gong Xin et al. [4] presented a feature matching method based on SURF. Firstly, the original image and box filter are convoluted to generate scale space, and the integral image is used to accelerate the convolution, and then the scale transformation is performed. The SURF operator is generated by geometric transformation factors such as rotation, and the positioning is completed by feature point coarse and fine matching. Compared with SIFT, the use of integral images greatly increases the speed of calculation.

In the second category method, Yu et al.[5] applied BP (back propagation) neural network to substation instrument positioning. Preprocessed samples and original image for histogram, binarization, edge detection, segmentation, normalization, etc. After processing, the BP neural network is trained, and input the image into the network for identification to complete the positioning. Wang Yuru[6] designed a channel-oriented packet convolution module and corresponding classification network CWGCNet (channel wise group convolutional network). The above neural network positioning devices do not include most of the substations. And they hardly analyzed the influence of network structure parameters on the recognition accuracy.

In the first method, the SURF matching has a large error, which causes the calculation of the homography matrix to be incorrect, and leads to the failure of the instrument positioning. The neural network has a learning function, and only needs to input different templates and results into the network, and the network learns to recognize the image and complete the positioning. Compared with the first method, the neural network has the highest positioning accuracy and the lowest mislocation rate.

3. The Proposed Method

3.1. Convolutional Neural Network

The CNN is a hierarchical model, which mainly includes an input layer, a convolution layer, a pooling layer, a fully connected layer, and an output layer.

The function of the input layer is to receive images, and the size of the input layer is in keeping with the size of the input images. Since the feature extraction of CNN is robust, so the input images do not be preprocessed under normal circumstances.

Features will be extracted by using convolution operations in the convolution layer. The more convolutional layers, the stronger ability to express features. However, the more model parameters, the more complexity of model training and predictive computation. In the convolutional layer, the feature image of the previous layer is convolved with a learnable kernel, and the result of the convolution forms the neurons of this layer through the output of the activation function, thus constituting the feature images. Normally, the convolutional layer is calculated as follows [7]:

$$X^l_i = f \left( \sum_{j \in \mathcal{M}_j} X^{l-1}_{i} * kernel^l_{ij} + B^l \right)$$

(1)

$l$ is the $l$ layer, $kernel$ is the convolution kernel, $B^l$ is the offset, $\mathcal{M}_j$ is a selection of the feature map, and each layer has a unique offset. $f(x)$ Represents a nonlinear activation function, commonly used sigmoid function and tanh function.
The pooling layer is usually after the convolutional layer, and its function is to reduce the size of the feature image, making the feature have a certain spatial invariance, while reducing the amount of calculation and preventing over-fitting. Common pooling operations have average pooling and maximum pooling.

The fully connected layer is located after feature extraction, connecting all neurons of the previous layer with each neuron of the current layer. The fully connected layer acts as a "classifier" that maps features to the sample mark space. If the \( l \) layer is a fully connected layer and the previous layer is also a fully connected layer, then the \( l \) layer of feature vectors is calculated as:

\[
x^l = f \left( w^l x^{l-1} + b^l \right)
\]

The output layer generally uses a linear full connection, and now the most common use is ‘Softmax regression’ that generates a prediction vector of an image category, \( y = (y_1, ..., y_M)^T \). And \( M \) indicates the number of categories, \( y_1 \) expresses as follows, \( w_i^l \) indicates regression weight for Softmax, \( i = 1, ..., M \).

\[
y_i = \frac{e^{-w_i^l x^{l-1}}}{\sum_{j=1}^{M} e^{-w_j^l x^{l-1}}} \tag{3}
\]

The training of network parameters generally uses the Backpropagation Algorithm (BP). The algorithm estimates the error of the layer above the output layer according to the error of the output layer, and recursively forwards layer by layer, and calculates the error of each layer in the reverse direction. More details were described in the article [8].

3.2. CNN Network Parameter Setting and Instrument Detecting

The instrument positioning of substation is divided into two processes. One is to construct and train CNN parameters, and the other is to detect the instrument in the input images.

In this paper, 4 categories of instruments were selected to verify the performance of the proposed algorithm, as shown in Figure 1. The designed substation instrument position convolutional neural network consists of 6 layers, including 2 layers of convolution layer, 2 layers of pooling layer, and 2 layers of fully connected layers. The first layer of convolutional layer uses 64 convolution kernels, the size is 5×5 pixel; the second layer of convolutional layer uses 32 convolution kernels, the size is 5×5 pixel; the first fully connected layer uses 512 Neurons; the second fully connected layer uses 7 neurons, and 7 neurons correspond to 7 types of instruments (Monitors, switches, oil level gauges, temperature controllers, and no instrumentation images). The network optimization template function is selected as cross entropy, and the optimization method adopts Adam (Adaptive Moment Estimation) algorithm. The original image size is 1920×1080 pixels, and the sub-image containing the instrument area is manually intercepted from the original image, and the size is 400×400 pixels as a training sample. In order to improve the training speed, the 400×400 pixel image is reduced to 40×40 pixel as an input image of the trainer.
After the network parameters were obtained, it is necessary to identify and locate the position of the instrument on the image. First, the original image is reduced from 1920×1080 pixel to 200×100 pixel; then, candidate window of target size are set. There are four possible target size windows: 40×40 pixel, 60×60 pixel, 80×80 pixel, and 100×100 pixel. Sliding the four windows on the images, and the obtained sub-image is then scaled to 40×40 pixels and input into the trained CNN to obtain the output. In the output vectors, if there is a number greater than 0.99, it means that there is an instrument in the current original image that needs to be located, and the candidate positioning area is the position where the target is located.

4. Experiment and the Result
In order to test the performance of the CNN in substation instrument positioning, the common substation instruments shown in Figure 1 are taken as samples. Samples were taken in the morning, noon, afternoon, and evening; the samples contained clear and blurred images, with and without instruments. Define the following variables: n1 is the number of training samples, n2 is the number of test samples; η is the ratio of the number of training samples to the total sample size, $\eta = \frac{n_1}{N}$. $\zeta$ is the correct rate for each types of instruments,

$$\zeta = \frac{\text{Each type of meter correctly identifies the quantity by itself}}{\text{Number of instruments tested by each type}}.$$  
$\chi$ is the false recognition rate,

$$\chi = \frac{\text{The number of images found the target on the image without the target}}{\text{Total number of images without the target}}.$$

4.1. Influence of Different Training Sample Numbers on Recognition Accuracy
In order to test the influence of the proportion of different training samples on the total number of samples on the recognition accuracy. Three sets of experiments were designed and the results are shown in Table 1. The parameter settings in the network describe in Table 1. It can be seen from Table 1 that with the increase of the proportion of training samples and training samples in the total sample, the accuracy of various types of instrument identification in substation is increasing, and the false location rate of the instrument is decreasing. We used Number to represent the number of training images, and used Rate to represent the recognition accuracy.
Table 1. Influence of different training sample numbers on recognition accuracy.

| Instrument names       | Experiment 1 (n1=707; n2=6660; η=10%) | Experiment 2 (n1=3330; n2=2000; η=50%) | Experiment 3 (n1=4662; n2=2000; η=70%) |
|------------------------|---------------------------------------|----------------------------------------|-----------------------------------------|
|                        | Number | Rate | Number | Rate | Number | Rate | Number | Rate |
| Monitor                | 146    | 74%  | 835    | 80%  | 1226   | 86%  |            |      |
| Temperature controller | 80     | 46%  | 375    | 52%  | 529    | 67%  |            |      |
| Oil level gauge        | 243    | 68%  | 1332   | 70%  | 1743   | 75%  |            |      |
| Feeder disconnector    | 125    | 63%  | 509    | 67%  | 725    | 71%  |            |      |
| Average accuracy       | 62.7%  |      | 67.2%  |      | 74.7%  |      |            |      |
| False location rate    | 22%    |      | 19%    |      | 16%    |      |            |      |

4.2. Influence of Convolution Kernel Size on Recognition Accuracy

Under the premise of keeping the CNN structure, we change the convolution kernel size (respectively set to 3×3 and 7×7). The number of training samples is 929, including monitor samples is 211, oil level is 410, switches is 120, thermostats is 80, and blurred images is 108. By testing 500 samples, the experimental results are shown in Table 2. It can be seen from Table 1 that the 3×3 convolution kernel has the highest total accuracy, which indicates that the 3×3 size has the ability to distinguish the characteristics of different instrument categories. At the same time, because the 3×3 convolution kernel is the smallest, its computational complexity is also minimal.

Table 2. Influence of different convolution kernel size on recognition accuracy.

| Instrument names       | Convolution kernel size |
|------------------------|-------------------------|
|                        | 3×3 | 5×5 | 7×7 |
| Monitor                | 81% | 74% | 79% |
| Thermostat             | 52% | 46% | 51% |
| Oil level gauge        | 69% | 68% | 70% |
| Switch                 | 68% | 63% | 65% |
| Average accuracy       | 67.5% | 62.7% | 66.2% |
| False location rate    | 24% | 22% | 26% |

4.3. The Effect of the Number of Convolution Kernels on Recognition Accuracy

The experiment is carried out with a 3×3 convolution kernel. Under the condition that the network structure and parameters are unchanged, the number of convolution kernels is changed, and the influence of the number of convolution kernels on the recognition accuracy is tested. Respectively defining c1 and c2 is the number of first and second convolutional layer convolution kernels. The experimental results are shown in Table 3. It can be seen that when the number of convolution kernels increases, the total recognition accuracy increases, but the increase is not large. When the convolution kernel is increased from 32 and 16 to 128 and 64 respectively, the average recognition rate is only increased by 1%. The false location rate is also less affected by the number of convolution kernels, and the floating does not exceed 3%.

Table 3. Influence of different convolution kernels number on recognition accuracy.

| Instrument names       | Convolution kernels numbers |
|------------------------|-----------------------------|
|                        | c1=128,c2=64 | c1=64,c2=32 | c1=32,c2=16 |
| Monitor                | 83% | 81% | 81% |
| Thermostat             | 50% | 52% | 49% |
| Oil level gauge        | 71% | 69% | 67% |
| Switch                 | 70% | 68% | 69% |
| Average accuracy       | 68.5% | 67.5% | 66.5% |
| False location rate    | 26% | 24% | 27% |
4.4. Comparison with Traditional Methods

In order to compare the proposed method with the traditional template matching method, and the traditional method based on matching feature points, a comparative experiment was designed. The comparison test was carried out by SURF based algorithm. The algorithm parameters of this paper are the same as those in experiment 1 in Table 1. The experimental results are shown in Table 4. It can be seen from Table 4 that the method based on convolutional neural network has the highest positioning accuracy and the lowest false location rate compared with the traditional template positioning method.

| Instrument names   | SURF matching | Method of this paper |
|--------------------|---------------|----------------------|
| Monitor            | 46%           | 86%                  |
| Thermostat         | 93%           | 67%                  |
| Oil level gauge    | 15%           | 75%                  |
| Switch             | 37%           | 71%                  |
| Average recognition accuracy | 69%         | 74.7%               |
| False location rate. | 31.6%     | 16%                  |

In summary, for the substation instrument positioning problems, when the convolutional neural network consists of two layers of convolutional layer, two layers of pooling layer, and two layers of fully connected layers, the number of convolutional layers and the size of the convolution kernel are accurate for instrument identification. The rate impact is small, and the number of training samples and the proportion of training samples to the total sample have a greater impact on instrument identification.

5. Conclusion

The core function of the substation inspection robot is that it can automatically identify various instruments instead of people. The higher the recognition accuracy, the greater the value of the robot. Accurate positioning of the instruments on the image is a prerequisite for instrument reading or status determination. The all-weather working characteristics of the inspection robots brings many challenges to the instrument positioning, especially the adaptability of the algorithm to various working conditions. The CNN takes the images as input when performing target recognition, and does not need to be preprocessed; the local field and weight sharing technology of the simulated biological neural network reduces the network complexity and reduces the number of optimization parameters; the pooling technology enhances the recognition robustness. The CNN learns the characteristics of each instrument autonomously through a large number of sample training. Compared with the traditional positioning method, the CNN instrument positioning method has no dependence on the target geometry, adopting a set of training algorithms for all instruments (no need to design different algorithms for different instruments), and has more weather conditions. The characteristics of strong adaptability.

6. Acknowledgement

This work was partially supported by the Sichuan Science and Technology Program China (2018GZ0385), Sichuan education department Program China (17ZB0095), Talent Import Fund of Chengdu University of Information Technology China (KYTZ201633) and the Fundamental Research Funds for the Central Universities (201810621091).

7. References

[1] Zhu B L, Guo L and Wu Q W 2017 Pointer instrument intelligent reading method based on ORB and improved Hough transform J. Instrument technology and sensors, 1, 29-34.
[2] Rublee E, Rabaud V, Konolige K, et al. 2011 ORB: an efficient alternative to SIFT or SURF C. Proceedings of the IEEE International Conference on Computer Vision. Barcelona, Spain.
[3] Fang H, Ming Z Q, Zhou Y F, et al. 2013 Instrument identification algorithm suitable for substation inspection robot. *Automation and instrumentation, 5*:10-14.

[4] Gong X, Meng L F and Zhang T, et al. 2017 SURF-based substation switches separation monitoring and identification system. *Instrument technology and sensors, 8*, 56-58.

[5] Yu H, Wang L K and Wang K. 2012. Application of Image Processing Technology and Neural Network in Power Equipment Identification. *J. Information and Computer (Theoretical Edition), 10*.

[6] Wang Y R. 2018 Substation equipment identification based on deep convolutional neural network. *D. Shenyang Agricultural University*.

[7] Xu S S, Liu Y G and Xu Sheng. 2013 Wood defect identification based on convolutional neural network. *J. Journal of Shandong University (Engineering Edition), 43(2)*:23-28.

[8] Chen T 2017 Research on Stereo Matching Technology Based on Convolutional Neural Network. *D. Zhejiang University*. 