

A Bolt Defect Recognition Algorithm Based on Attention Model

Zhijun LIN¹, Yingjie LIANG and Qineng JIANG

*Jiangmen Power Supply Bureau of Guangdong Power Grid Co., LTD
Jiangmen, China*

Abstract. As the largest number of fasteners on power distribution network, bolts are the cornerstone of ensuring the safety and reliability of power system. However, pin losing, nut losing, nut loosening, and rusting can cause damage to power system and even cause terrible accidents. In order to solve the problem that the large number of bolt defects causes traditional manual identification to be difficult and inefficient, this paper proposes a bolt defect identification algorithm based on attention models. The method in this paper improves the traditional deep residual network ResNet network, adds a channel attention mechanism to obtain key channel features, and uses random flipping, translation and other data augment methods to expand the bolt defect dataset. The experimental results show that compared with the traditional model, the improved model can more accurately identify different types of bolt defect images, and the mAP on the testing set reaches 85.9%, which verifies the feasibility and reliability of the ATT-ResNet50 model in bolt defect recognition. The method proposed in this paper has high recognition accuracy, realizes the intelligent recognition of common bolt defects.

Keywords. Attention Mechanism, Deep Residual Network, Data Augment, Bolt Defect Recognition

1. Introduction

As the most basic fastener of power distribution network, bolts produce an essential effect on the deployment of power grids. Affected by a variety of external factors such as the environment, the connecting bolts of the components are caused to produce defects, which in turn will affect the safe connection between the various components of the power distribution network, and cause safety risks to the power system [1,2]. Pin losing, nut losing, nut loosening and rusting are the four most common types of defects in bolts. These defects may cause loose connection of power system components, which not only increases the loss of power system, but also presents the safety hazards of components falling even terrible grid accident [3].

In recent years, for the sake of raising the inspection efficiency of power distribution network, drone inspection technology has been widely used. Regarding the external defects of power grid equipment, drones with data acquisition devices can quickly obtain a large number of pictures and then corresponding defects will be found

¹ Corresponding Author, Zhijun LIN, Jiangmen Power Supply Bureau of Guangdong Power Grid Co., LTD Jiangmen China; E-mail: 7648625@qq.com. The authors would like to thank the anonymous reviewers for their comments on this paper. This research was supported by the science and technology project of China Southern Power Grid Company Limited under Grant No. GDKJXM20198068 (030700KK52190190)

manually. In this process, faced with a large amount of image data, even if it is annotated by professionals with rich experience, its efficiency is very low. For this reason, it has great significance to study an accurate and effective defect recognition technique for common bolt defects.

Traditional power component and defect recognition methods mainly focus on the design of manual features. For example, after image preprocessing with the help of image enhancement and denoising methods, Harr features, moment invariants, color space and other features are combined with support vector machines, level combined with Adaboost and other algorithms to realize the identification of anti-vibration hammers, insulators and other electrical components and corresponding defects [4,5]. In addition to requiring a wealth of professional knowledge support, such methods often only work for a specific category and have poor scalability. Nowadays, deep learning technology has made long-term development, it has been widely employed in all aspects of life. In the meantime, the recognition algorithm based on deep convolutional neural networks has also solved a lot of defect recognition problems and achieved many results [6]. However, its research on the identification of bolt defects is relatively insufficient. The paper [7] uses the traditional object detection algorithm to detect the bolts of the transmission line, by constructing the transmission line inspection image dataset, extracting the HOG features of the bolts, using the SVM classifier to classify, realizes the identification of the bolts in the inspection image, but this method only simple identification of the bolt target is carried out, for the more important bolt defect identification work, further in-depth analysis and research are lacking. In reference to the problem of sample imbalance, the paper [8] introduced auxiliary data and proposed a RetinaNet-based method for identifying missing and loosening bolts of power system, which achieved better performance results, but the types of bolt defects were relatively single.

In view of the problems above, this paper adds an attention module to the ResNet so as to weight different channels more effectively and obtain more key feature channels. At the same time, considering that it is short of training sample of bolt defect dataset which will cause a risk of residual network overfitting. This paper uses a data augment strategy to expand the training dataset, and sets up experiments to discuss the impact of increased data on the accuracy of defects. The results show that adding an appropriate amount of data can notably promote the recognition accuracy of the model.

2. Deep Convolutional Neural Network Model and Attention Mechanism

2.1 Deep Convolutional Neural Network

In the 2012 ILSVRC image classification competition, KRIZHEVSKY et al. proposed AlexNet [9], a classic deep convolutional neural network model. Compared with the shallow convolutional neural network model, AlexNet obtains a significant improvement in performance, and the deep network model has a greater advantage than the shallow network model. The VGG16 [10] network and GoogleNet [11] network constantly refresh the accuracy of the ILSVRC competition. From LeNet [12], AlexNet to VGG16 and GoogleNet, the number of layers of convolutional neural networks is constantly increasing. With the deepening of the number of network layers, the amount of data and the amount of computing are also increasing sharply. The mathematical expression of the convolution module is shown in Equation (1):

$$\begin{cases} y = \text{conv}(x, w) = (y(1), y(2), \dots, y(n - m + 1)) \in R^{n-m+1} \\ y(t) = \sum_{i=1}^m x(t + i - 1)w(i) \end{cases} \quad (1)$$

Among them, x is the input, w is the size of the convolution kernel and $w \in R^{s \times k}$, $x \in R^{n \times m}$, $t=1, 2, \dots, n-m+1$. The outstanding contribution of convolution lies in that it can cut down unnecessary weight connections, introduce sparse or partial links, and the weight sharing strategy brought about greatly reduces the amount of parameters, relatively increases the amount of data, so as to avoid overfitting [13]. What's more, owing to the translation invariance of the convolution, the characteristics learned possess topological robustness and symmetry.

2.2 ResNet Network

As one of the classic models in the field of deep learning, ResNet [14] has surpassed VGG in image classification problems and has become the basic feature extraction network in most visual fields. ResNet introduces the residual element, which makes it not only compress the parameters, but also add a direct channel in the network to further improve the ability of feature learning [15]. Different from the early network, its features such as fewer parameters, deep layers and excellent classification and recognition effects make it still one of the more classic and useful networks so far. The residual module of the network model based on the SE module is shown in Figure 1.

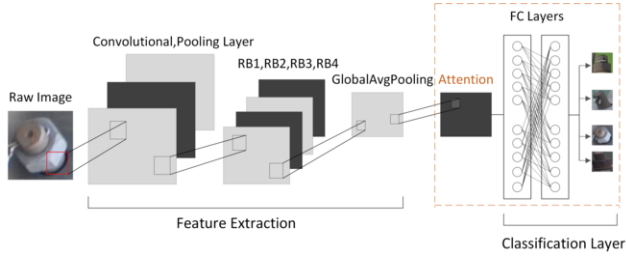


Figure 1 The framework of deep residual network for bolt defect recognition based on attention mechanism

2.3 Attention Mechanism

There are two major kinds of common attention mechanisms, one kind is spatial attention, such as spatial transformer network, the other is channel attention, for instance BAM [16] and SE module(Squeeze-and-Excitation Networks) [17]. In general, the key task of channel attention is to explore discriminative feature. As one of the classic channel attention, SE module is able to transform the response of the channel feature by finding the correlation between different channels [18]. Compared with the traditional neural network, this method cuts down the number of calculation to a great extent. The key purpose of spatial attention is to learn detailed features as much as possible. The STN(spatial transformer network) explicitly allows spatial operations on data and allows data to be processed to enhance the geometric invariance of the model. Besides, it is able to be inserted into any existing convolutional model and only a few modifications are required.

The advantage of attention mechanism is that it is able to distribute different weights to distinct parts of the input so as to help the network obtain as much key information as possible. What is more, it doesn't need additional calculations. Generally speaking, we can divide attention into hard attention and soft attention. Among them, soft attention tends to focus on spaces as well as channels. There are some difference between hard attention and soft attention, for example hard attention focuses more on stochastic prediction and tends to emphasizes the dynamic variations of the model. Consequently, it is hard to apply end-to-end ways to train hard attention and the most commonly used method is reinforcement learning.

In this study, for the sake of obtaining as many discriminant features as possible, we apply the channel attention mechanism. Among them, the SE module is a classic channel attention method. It finds the relationship between channels to freely Adapt to change the channel characteristic response. Compare with the traditional neural network, the core idea of SE module lies in learning the feature weight according to loss through the network, so as to train the model in a way that the weight of effective feature map is large when the weight of ineffective feature map is small, so as to reach better results. Although the process of embedding SE module into some original classification networks inevitably brings a small amount of parameters and computation, it is acceptable.

By means of modeling the interdependencies among the feature channels, SE module is able to ameliorate the representation of the network [19]. Furthermore, Se module can be divided into three parts: squeeze, excitation, and relabeling [20]. First, perform squeeze operation on the feature map calculated by convolution to get the global feature of each channel, then perform extraction operation on the global feature to get the relationship and weight of each channel, and finally use the channel weight to multiply the original image so that the final feature distribution of the feature map is obtained. The schematic diagram of the SE block is shown in Figure 2.

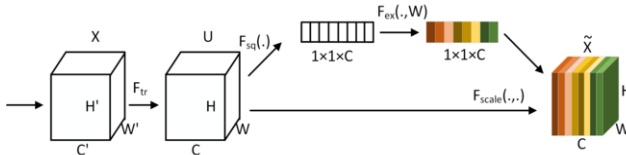


Figure 2 The schematic diagram of the SE block

During squeeze operation, global average pooling is adapted to average all information on a channel to obtain the global features on the channel, which solves the problem of small receptive fields in the CNN network. The calculation method is shown in Equation (2):

$$z_c = F_{sq}(u_c) = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H u_c(i, j) \quad (2)$$

Among them, $u_c(i, j)$ represents a pixel in the image, W and H represent the width and height of the image respectively. Moreover, squeeze operation sums all the pixel values and takes the average value.

Excitation operation requires the relationship of information in each channel, which represents as Equation (3). It takes two fully connected bottleneck structure, among them, $W_1 \in R^{\frac{C}{r} \times C}$, $W_2 \in R^{\frac{C}{r} \times C}$, r is a hyperparameter, σ and δ are two activation

functions. After calculation, the image features compressed by Equation (3) can be extracted.

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1, z)) \quad (3)$$

Finally, use the learned weight parameters to multiply each channel feature calculated by the original convolutional network to calculate the output of SENet, as shown in Equation (4):

$$x_c = F_{scale}(u_c, s) = s_c \cdot u_c \quad (4)$$

Among them, u_c represents the feature image of each channel calculated by the formula and s_c represents the weight of the channel, which is multiplied to obtain the fused image information.

2.4 Application of attention mechanism in the model

The attention mechanism can improve the network's ability to extract regions of interest, and it can also improve the accuracy of network recognition. This paper proposes to add an attention module to the ResNet classic network, and to improve the model parameters, and finally form a network structure suitable for bolt defect recognition task ATT-ResNet50. The core of ResNet50 based on the attention mechanism is the attention layer framed by the dashed line inside the model. The network with the attention module has more powerful feature extraction capabilities. In the meantime, the depth advantage of the ResNet network layer makes the effect more significant [21].

3. Experimental results and analysis

3.1 Dataset

This paper uses the bolt defect dataset to evaluate the proposed model. The bolt defect pictures are all intercepted from samples taken by drone line inspections. After data cleaning and careful data selection, the data set that can be used in this experiment is sorted out. There are four types of data samples, a total of 2000, and the training set and testing set are divided according to the ratio of 7:3. Among them, there are 500 pictures of pin losing, nut losing, nut loosening and rusting respectively. These four kinds of pictures are put respectively in distinct folders and labeled 0-3. The experimental dataset without data augment is shown in Figure 3.



Figure 3 Bolt defect dataset

3.2 Experimental details

The experiment in this paper is implemented using the PyTorch framework, and a single NVIDIA Geforce GTX Titan X GPU is used to train and test the method in this paper. First, build the AlexNet, VGG and ResNet networks, and then load the pre-trained AlexNet, VGG and ResNet50 model weight files, freeze the feature extraction layer, and remove the global average pooling layer and FC layer of the original network. The channel attention module layer is embedded in the model to make the features obtain different weights, and then a new global average pooling layer and FC layer are added to facilitate subsequent model training [22]. Before training, the learning rate is set to 0.001, the dropout rate is set to 0.5. Moreover, we apply SGD [23] algorithm to update the parameters of model.

3.3 Influence and analysis of attention mechanism on experimental results

For the sake of fully validating the performance of the improved model method proposed in this study, the bolt defect dataset is respectively carried out on AlexNet, VGG, ResNet and the above-mentioned network with the attention mechanism to perform identification experiments, and the results are recorded in table 1, which shows the precision comparison between the network models. As we can see from the table, compared to AlexNet and VGG16, ResNet50 is the most effective basic network, and it can achieve a mAP of 82.2% on the bolt defect dataset. After introducing attention mechanism, the recognition effects of these three models have achieved obvious results, and the improvement is the most obvious on the ResNet50 model, not only the mAP is maximized, but the AP of each category is higher, which proves that adding channel attention can effectively ameliorate the recognition effect of the model on the bolt defect dataset.

Table 1 Comparison of bolt defect recognition on different models

models	0	1	2	3	mAP
AlexNet	76.3	76.7	74.4	80.5	77.0
ATT-AlexNet	77.1	76.1	74.8	81.0	77.3
VGG16	78.3	77.9	76.0	81.7	78.5
ATT-VGG16	80.5	79.2	76.7	82.3	79.7
ResNet50	82.6	83.0	78.4	84.8	82.2
ATT-ResNet50	84.4	85.3	80.3	85.5	83.9

Table 2 The relationship between AP and the number of training set

Number of training set	0	1	2	3	mAP
1*1400	84.4	85.3	80.3	85.5	83.9
2*1400	85.6	86.6	81.7	86.5	85.1
3*1400	85.9	87.1	82.3	87.1	85.6
4*1400	86.2	87.3	82.2	87.8	85.9
5*1400	85.4	86.6	81.5	86.4	85.0
6*1400	84.2	85.0	79.8	85.2	83.6

3.4 Quantitative experiments and analysis of data augment

In order to analyze the influence of data augment on the test results, under the original parameter settings, for the ATT-ResNet50 model, different multiples of data samples are added to the original bolt defect dataset for model training, and the same batch of testing set is used to perform model testing. The model is tested and the results

obtained are shown in the table below. From the results in Table 2, it can be seen that after adding different multiples of training data, the model can learn more features during the training process, and there is a significant improvement on mAP from 83.9% to 85.9%. Therefore, when the number of training samples is insufficient to support deep learning defect recognition, the generalization ability of the model can be improved by adding the number of training samples. Furthermore, if the number of training samples reaches four times the original training samples, the mAP reaches the maximum value of 85.9%. However, if the number of training samples is further increased, when the number of training samples reaches five times or more, the mAP will drop sharply.

4. Conclusion and Future Work

Considering that traditional manual identification of common bolt defects on power fittings is consuming with long time as well as labor, and misdetection often occurs, this paper puts forward a bolt defect identification algorithm based on attention model. By improving the traditional deep residual network ResNet50 and adding a channel attention mechanism to obtain key channel features, the effective improvement of defect recognition rate has been realized, which provides a theoretical basis for intelligent defect recognition in the future. In addition, considering that there are fewer samples of bolt defects in reality and the high cost of collection, this paper uses data augment to assist the training of the model, and uses experiments to verify that an appropriate raise of training samples is conducive to improve the generalization ability.

In the future, we would like to expand the current bolt defect data samples and types, and take more effective ways to further improve the model.

References

- [1] Janos Toth, and Adelana Gilpin-Jackson. (2010). Smart view for a smart grid : Unmanned Aerial Vehicles for transmission lines. *Applied Robotics for the Power Industry (CARPI)*, 2010 1st International Conference on IEEE.
- [2] Van Nhan Nguyen, Robert Jenssen, and Davide Roverso. (2018). Automatic autonomous vision-based power line inspection: A review of current status and the potential role of deep learning. *International Journal of Electrical Power & Energy Systems* 99. JUL. (pp. 107-120).
- [3] Chuang Deng, Shengwei Wang, Zhi Huang and Zhongfu Tan. (2014). Unmanned aerial vehicles for power line inspection: a cooperative way in platforms and communications. *Journal of Communications*. (pp. 687-692).
- [4] Xian Tao, Dapeng Zhang, Zihao Wang, Xilong Liu, Hongyan Zhang, De Xu. (2018). Detection of power line insulator defects using aerial images analyzed with convolutional neural networks[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018, 50(4): 1486-1498.
- [5] Changfu Xu, Bin Bo, Yang Liu, Fengbo Tao. (2018). Detection method of insulator based on single shot multibox detector[C]//*Journal of Physics: Conference Series*. IOP Publishing, 2018, 1069(1): 012183.
- [6] Xiongwei Wu, Doyen Sahoo and Steven C.H.Hoi. Recent advances in deep learning for object detection[J]. *Neurocomputing*, 2020, 396: 39-64.
- [7] Min Feng, Wang Luo, Lei Yu, Pei Zhang, Xiaolong Hao, Qiang Fan, Qiwei Peng, Tianbing Zhang and Lingling Cao. (2018). A bolt detection method for pictures captured from an unmanned aerial vehicle in power transmission line inspection[J]. *Journal of Electric Power Science and Technology*, 2018, 33(4): 135-140.

- [8] Kai Wang, Jian Wang, Gang Liu, Wenqing Zhou and Zhuoyang He. (2019). RetinaNet algorithm based on auxiliary data for intelligent identification on pin defects[J]. *Guangdong Electric Power*, 2019, 32(9): 41-48
- [9] Alex Krizhevsky, I Sutskever and G Hinton. (2012). ImageNet classification with deep convolutional neural networks[C]//NIPS, 2012.
- [10] Karen Simonyan and Andrew Zisserman. (2015). Very deep convolutional networks for large-scale image recognition[J]. *arXiv Preprint*, 2015: arXiv: 1409. 1556.
- [11] Christian Szegedy, Wei Liu, Yangqing Jia and Pierre Sermanet. (2015). Going deeper with convolutions[C]// *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Washington, DC: IEEE Computer Society, 2015: 1-8.
- [12] SooBum Kim, JiHoon Lee, SeungYeon You, SungWook Kim and SungChun Kim. (2006). Power-aware Pat Selection Scheme for AOMDV[J]. *Brain Korea21 Project I*, 2006.
- [13] Bin Cheng. (2019). High resolution image classification of urban areas based on convolution neural network[C]//2019 4th International Conference on Mechanical, Control and Computer Engineering (ICMCCE). IEEE, 2019: 417-4174.
- [14] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun. (2016). Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), June 27-39, Las Vegas, NV, USA, New York: IEEE, 2016: 770-778
- [15] Xiaoxu Li, Jijie Wu, Dongliang Chang, Weifeng Huang, Zhanyu Ma, Jie Cao. (2019). Mixed Attention Mechanism for Small-Sample Fine-grained Image Classification[C]//2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2019: 80-85.
- [16] Jongchan Park, Sanghyun Woo, Joon-Young Lee, In So Kweon. (2018). Bam: Bottleneck attention module[J]. *arXiv preprint arXiv:1807.06514*, 2018.
- [17] Jie Hu, Li Shen, Gang Sun and Samuel Albanie. (2018). Squeeze-and-excitation networks[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 7132-7141.
- [18] Zhen Chen, Maoyong Cao, Peng Ji and Fengying Ma. (2021). Research on Crop Disease Classification Algorithm Based on Mixed Attention Mechanism[C]//*Journal of Physics: Conference Series*. IOP Publishing, 2021, 1961(1): 012048.
- [19] Qiang Chen, Li Liu, Rui Han, Jiaying Qian, Donglian Qi. (2019). Image identification method on high speed railway contact network based on YOLO v3 and SENet[C]//2019 Chinese Control Conference (CCC). IEEE, 2019: 8772-8777.
- [20] Guihui Shi, Jiezhong Huang, Junhua Zhang, Guoqin Tan, Gaoli Sang. (2021). Combined Channel and Spatial Attention for YOLOv5 during Target Detection[C]//2021 IEEE 2nd International Conference on Pattern Recognition and Machine Learning (PRML). IEEE, 2021: 78-85.
- [21] Xianghan Wang, Jie Jiang, Yanming Guo, Yingying Gao, Jun Lei, Lai Kang, Yingmei Wei. (2019). FACPM: Crop Hand Area with Attention[C]//2019 5th International Conference on Big Data and Information Analytics (BigDIA). IEEE, 2019: 139-143.
- [22] Matthew D. Zeiler. (2012). Adadelta: an adaptive learning rate method[J]. *arXiv preprint arXiv:1212.5701*, 2012.
- [23] Bottou L. (2010). Large-scale machine learning with stochastic gradient descent[M]//*Proceedings of COMPSTAT'2010*. Physica-Verlag HD, 2010: 177-186.