Detection of Common Foreign Objects on Power Grid Lines Based on Faster R-CNN Algorithm and Data Augmentation Method

Guangxin Zu¹, Guoliang Wu¹, Chong Zhang²,* and Xing He²

¹Electric Power Research Institute of State Grid Heilongjiang Electric Power Co., LTD, Harbin, China
²School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China

*Corresponding author email: zc2640@sjtu.edu.cn

Abstract. Foreign objects hanging on the electrical equipment are big threats to the safety of the power grid. In this paper, a high-precision object detection model designed for common foreign objects in power grid based on Faster R-CNN algorithm is proposed. The data augmentation methods such as rotation, adjusting brightness and colour saturation, adding Gaussian noise are carried out to expand the data in the training set. The accuracy rate and the recall rate of the model reach 93.09% and 94.33% respectively. The object detection model trained by the training set with data augmentation has achieved better detection results.

Keywords: Foreign objects; Faster R-CNN; Data augmentation.

1. Introduction

That the electrical equipment such as wires, poles and towers are wrapped by foreign objects are an important factor leading to the contact short circuit between lines. In the traditional inefficient manual inspection mode, power workers have a large workload considering that the scale of China's transmission and distribution network. In recent years, with the quick development of computer vision technology, using target detection technology to deal with UAV patrol photos can achieve automatic detection of power grid line fault. It is an inevitable trend to improve the intelligence of line inspection. The object detection model based on Faster R-CNN algorithm is selected to detect the common foreign objects in the power grid, so as to improve the intelligent level of the inspection process. This paper designs image data augmentation and multi-scale adjustment method to improve the detection effect of the object detection model.

2. Object Detection Model Based on Faster R-CNN Algorithm

The object detection algorithms mainly include two types: two-stage methods and one-stage methods. The two-stage methods mainly include Region-Convolutional Neural Network (R-CNN)[1] and Fast R-CNN[2], Faster R-CNN[3], R-FCN[4], Mask R-CNN[5] and other optimization algorithms. One-stage methods include You Only Look Once (YOLO)[6] and Single Shot Multi-Box Detector (SSD)[7]. In reference[8] , Faster R-CNN is used in the identification and defect detection of components on the transmission lines such as grading rings and shock hammers. In reference [9], a multi-target detection and positioning model based on improved Faster R-CNN is proposed to identify and detect electrical equipment such as overhead lines, towers, insulators and fittings. In reference [10], Faster R-CNN technology is used to identify the self-explosion defects of insulators in UAV images.
Compared with one-stage methods, the two-stage methods can detect small targets with high precision but slow speed. Considering the uncertainty of size of foreign objects in power grid, in order to ensure a good detection effect, we should not only pay attention to the detection speed, but also ensure the detection rate of foreign objects. Faster R-CNN can achieve satisfactory detection results for small objects, so this paper chooses the object detection model based on Faster R-CNN algorithm.

The Faster R-CNN algorithm framework model is shown in figure 1. The original image is first extracted by convolutional neural network to get the feature map. Then the feature map enters the region proposal networks (RPN), a large number of anchor boxes are generated. After filtering the anchor boxes, one branch judges whether the anchor boxes belong to objects or background. At the same time, the other branch modifies the anchor boxes by using smooth loss function to form more accurate proposals. Region of Interest (ROI) pooling layer synthesizes the proposals generated by module of RPN and the feature map generated by CNN feature extraction network to get a fixed size proposal feature map. Finally, the proposal feature map generated by ROI layer enters the full connection layer network, and outputs the exact location of the bounding boxes and the specific categories of the objects through the border regression layer and classification layer.

![Figure 1. Model of Faster R-CNN algorithm.](image)

2.1. **CNN**

CNN carries out the convolution of image data with a certain scale of convolution kernel, which can well extract the information of edge, texture and other features of objects in the image. The basic structure of CNN includes input layer, convolutional layer, pooling layer, fully connected layer and output layer.

2.2. **RPN**

RPN uses sliding window to extract multiple candidate areas from input image, and performs foreground/background discrimination for each candidate area and calculates position coordinates. The structure of RPN is shown in figure 2. In the sliding process, the mapping point of sliding window center in original pixel space is called anchor, $k$ candidates are generated with this anchor as the center and put through two fully connected layers to get the results of object scores and the object bounding boxes regression.

![Figure 2. RPN network structure.](image)

2.2.1. **Classification layer.** In the branch of classification layer, the foreground / background of proposals is classified by full connection layer and softmax function. Softmax function can compress k-dimension real number vector into a new dimension real number vector with a range of (0-1) to
normalize the output value to probability value in multi classification problems. The function form is shown in equation (1).

\[
\sigma(z)_j = e^{z_j} \left( \sum_{k=1}^{K} e^{z_k} \right)^{-1}, (j = 1, \ldots, K)
\]  

(1)

where \( z_j \) is the value input into the neuron \( j \) in the output layer, and the result output from softmax function represents the probability belonging to the class \( z_j \).

2.2.2. **Regression layer.** In the process of sliding window scanning, anchor boxes are generated at each anchor point. The intersection over union (IoU) between the anchor boxes and the ground truth of the object is used as an indicator to measure and mark the anchor boxes\(^{[12]}\). The calculation formula of IOU is shown in equation (2).

\[
IoU = \frac{(A \cap B)}{(A \cup B)}
\]  

(2)

2.2.3. **The loss function of RPN.** \( P_i \) is the probability that is the ith anchor box contain the object. If \( p_i^* \) is a positive sample, its value is 1 and if \( p_i^* \) is a negative sample, its value is 0. The loss function of the classification layer is shown in equation (3).

\[
L_{cls}(p_i, p_i^*) = -\log[p_i^* p_i + (1-p_i^*) (1-p_i)]
\]  

(3)

The regression layer loss function is shown in equation (4).

\[
L_{reg}(t_i, t_i^*) = R(t_i - t_i^*)
\]  

(4)

Where \( R \) function is smoothL1 function, its definition is as shown in equation (5).

\[
smooth_{L1}(x) = \begin{cases} 
0.5x^2, & |x| < 1 \\
| x | - 0.5, & |x| \geq 1
\end{cases}
\]  

(5)

The total loss function of RPN is:

\[
L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i L_{reg}(t_i, t_i^*)
\]  

(6)

Where \( N_{cls} \), \( N_{reg} \) and \( \lambda \) can be regarded as normalized weights to balance classification function loss and regression function loss. Through training, the total loss function value is minimized to achieve good classification and regression effect.

3. **Training Methods for Detection of Common Foreign Objects on Power Grid Lines**

3.1. **Construction of Training Set**

3.1.1. **Data augmentation of training set.** The performance of the neural network model is closely related to the number of samples in the training set. With the increase of the amount of data in the training set, the parameters in each layer of neural network gradually converge to the optimal value. Because of the randomness of foreign objects on power lines, it is hard to collect and build a training set containing a large number of foreign objects image samples. Therefore, this paper uses image processing methods of rotation, changing brightness and colour saturation to realize the data augmentation and expand the existing training set. The image processing method is shown in table 1. By writing the automatic image processing algorithm program, the expansion of training set can be realized efficiently.
Table 1. Data enhancement method.

| rotation | brightness | colour saturation | mirror         |
|----------|------------|-------------------|----------------|
| 20°      | original image *0.6 | original image *0.5   | left to right |
| 160°     | original image *1.4 | original image *1.5   | top to bottom |

Gaussian noise refers to a kind of noise of which probability density function obeys Gaussian distribution (i.e. normal distribution). The colour image can be divided into three channels of RGB, and Gaussian noise is added to each channel, so that the added Gaussian noise will form colour noise points on the image. Adding Gaussian noise to the image is also an effective way to realize the data augmentation. The image before and after adding Gaussian noise are respectively shown in the figure 3.

![Image](image_url)

(a) The original image  (b) The image after adding Gaussian noise

Figure 3. The image before and after adding Gaussian noise.

After the data augmentation method is adopted, the number of image set is expanded ten times as much as the original.

3.1.2. Multi scale adjustment of pictures in training set. The diversity of images in training set is caused by the different sources, the different definition and size of image samples. Too large pictures increase the calculation of the model in the training process, too small pictures are not conducive to the extraction and learning of the target features. In order to improve the speed of training process and the adaptability and robustness of the network to different scales, the size of the image is adjusted in the data preprocessing stage, and the long side size of the image is adjusted according to 400,500,600 while maintaining the original scale of the image.

3.2. Selection of CNN Model
In this paper, GoogLenet[13] Inception V2 model is adopted. The sparse network structure is designed in Inception V2. In addition, the convolution kernels of different sizes, such as 1×1 and 3×3 are used for convolution in parallel, which not only increases the network width, but also enhances the adaptability of the network to scale. Using 1×1 convolution kernels for dimensionality reduction, using two continuous 3×3 convolution kernels to replace single 5×5 convolution kernels, and using continuous 1×N and N×1 convolution kernels instead of n×n convolution kernels on medium-sized feature graphs can greatly reduce the amount of parameters and reduce the degree of over fitting while ensuring almost the same expression effect. At the same time, BN regularization is used to accelerate the speed of training process of the network and improve the classification accuracy after convergence.

4. Case Studies
4.1. Construction of Training Set
Due to the diversity and uncertainty of foreign objects in power grid lines, there are many detection objects. In this paper, 150 pictures of bird's nest with different light intensity, shape and angle were selected for training. 180 pictures of balloon are selected, including balloons from different kinds in multiple scenes as training sets. 200 pictures of kite taken in different angles in various scenes are selected. By improving the diversity of training samples to train the object detection model, high
detection accuracy can also be obtained in the multi scene pictures including complex environment and background interference.

4.2. Training Process
The hardware configuration of this experiment includes a computer host (32g CPU) and an independent graphics card RTX 2080 Super (used to speed up the operation process of training neural network model). The software configuration is windows10, anaconda3, python3.6, CUDA tool kit 10.0, cudnn8.0, tensorflow-gpu1.9, etc.

This experiment uses the faster_rcnn_inception_v2 pre trained on the COCO data set, which can reduce the sample size of training set and improve the efficiency of the training process. The hyper-parameter of iterative training are set as follows:
1) step1-15000, learning-rate=0.0003;
2) step15001-20000, learning rate=0.00003;
3) step20001-25000, learning rate=0.000003.

The results of training loss function are shown in figure 4. With the number of iterations increasing from 0 to 25K, the results of object detection classification and regression loss function tend to zero.

![Classification loss function curve](image1)
![Localization loss function curve](image2)

Figure 4. The results of training loss function.

4.3. Test Results and Analysis
Batch processing 300 pictures including kites, balloons and bird’s nests, and then use them to test the performance of object detection model. Some test results are shown in figure 5.

![Bird's nest test results](image3)
![Kite test results](image4)
(c) Balloon test results

Figure 5. Some test results of test set data.

The precision and recall rate are used as the indexes to evaluate the performance of the object detection model\cite{14}, and the calculation formulas are shown in equation (7) and equation (8):

\[
\text{precision} = \frac{TP}{TP + FP} \quad (7)
\]

\[
\text{recall} = \frac{TP}{TP + FN} \quad (8)
\]

Where TP (true positions) refers to that positive samples are correctly identified as positive samples; TN (true negatives) refers to that negative samples are correctly identified as negative samples; FP (false positions) refers to that false positive samples; FN (false negatives) refers to that false negative samples.

In order to improve the detection effect, this paper uses data augmentation method to expand the data set. The test results of the object detection model trained by the training set before data enhancement are recorded and compared with the test results of the object detection model trained by the expanded training set. The test results are shown in the table 2.

| catalogue                      | TP   | FP   | FN   | precision   | recall  |
|--------------------------------|------|------|------|-------------|---------|
| before data augmentation       | 86   | 18   | 14   | 82.69%      | 86.0%   |
| After data augmentation        | 283  | 21   | 17   | 93.09%      | 94.33%  |

Table 2. Effect comparison of two object detection models.

As is seen from the test results that testing the object detection model based on Faster R-CNN trained using the original training set without data augmentation method, the precision rate of this model is 82.69%, and the recall rate is 86.0%. By contrast, the precision of the object detection model trained by the expanded training set is as high as 93.09%, the recall rate is as high as 94.33%, the detection effect has been significantly improved.

In order to compare the performance of the object detection model based on the Faster R-CNN, this paper uses the same expanded training set to train the object detection model based on the YOLOv3 algorithm. The software configuration is cuda-10.1,cudnn-10.1,visual studio 2017 community, opencv3.4, etc. the effect of the object detection model based on the YOLOv3 and Faster R-CNN algorithm on the test set is shown in table 3.

| catalogue | precision | recall | Speed (ms·sheet\(^{-1}\)) |
|-----------|-----------|--------|---------------------------|
| Faster R-CNN | 93.09%    | 94.33% | 148                       |
| YOLOv3    | 84.72%    | 84.34% | 91                        |

Table 3. Comparison of test results of two algorithms.

By analyzing the data in table 3, it can be known that although the detection speed of model based on YOLOv3 algorithm is faster, however, its detection effect is not as good as the object detection model based on Faster R-CNN algorithm, and the accuracy and recall rate of the former are 84.72% and 84.34% respectively, which are lower than the data corresponding to the latter. Because the hidden danger caused by the suspension of foreign objects should be minimized as far as possible. The precision and recall rate of the object detection model based on the YOLOv3 algorithm are lower than that based on
the Faster R-CNN algorithm, so the Faster R-CNN algorithm is more suitable for the task of power lines inspection.

5. Conclusion

In this paper, the object detection model is proposed for kites, balloons and bird’s nests, which are common foreign objects on power grid lines. The training set data is pre-processed by data augmentation and multi-scale adjustment method and input to the model based on Faster R-CNN algorithm. The precision and recall rate of the object detection model on test set data can reach 93.09% and 94.33% respectively, and the detection effect is satisfactory. In addition to the common types of kites, balloons and bird's nests, there are also the types of foreign objects without obvious shape and edge features, such as plastic bags, plastic films and textiles. Further research is needed on the construction of training sets for such foreign objects and the optimization of object detection model.

References

[1] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]// CVPR. IEEE, 2014
[2] Girshick, Ross. Fast R-CNN[J]. Computer Science, 2015.
[3] Ren, Shaoqing, He, Kaiming, Girshick, Ross, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 39(6):1137-1149.
[4] Dai J, Li Y, He K, et al. R-FCN: Object Detection via Region-based Fully Convolutional Networks[J]. 2016.
[5] He Kaiming, Gkioxari Georgia, Dollar Piotr,. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence:1-1.
[6] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[J]. 2015.
[7] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[C]// European Conference on Computer Vision. Springer International Publishing, 2016.
[8] TANG Yong, HAN Jun, WEI Wenli, et al. Deep learning in component identification and defect detection of transmission lines [J]. Electronic measurement technology, 2018.
[9] LIN Gang, WANG Bo, PENG Hui, et al. Multi-object detection of transmission line patrol image based on improved faster-rcnn [J]. Power automation equipment, 039 (005) : 213-218.
[10] GUO Tao, YANG Heng, SHI Lei, et al. Self-explosion defect identification of insulators based on Faster RCNN [J]. Electric porcelain arrester, 2019 (3).
[11] Shin H C, Roth H R, et al. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning[J]. IEEE Transactions on Medical Imaging, 2016, 35(5):1285-1298.
[12] H. SHENG, M. HUANG and J. YANG. Laboratory population statistics and management system optimization based on faster r-cnn and IOU [J]. Computer application, 2019, 39 (6): 1669-1674
[13] Pedro Lemos Ballester, Ricardo Matsumura Araujo. On the performance of GoogLeNet and AlexNet applied to sketches[C]// AAAI. AAAI Press, 2016.
[14] Y. ZHANG, Z. XU, D. FENG, et al. Research on fast region convolutional neural network target detection based on hard sample mining [J]. Journal of electronics and information, 2019, 41 (6)