An insulator inspection method based on deep learning applicable to multi-scale and occlusion conditions

Junpeng Wu | Shaobo Tang | Xianglei Li

School of Electrical Engineering, Northeast Electric Power University, Jilin, People’s Republic of China

Correspondence
Shaobo Tang, School of Electrical Engineering, Northeast Electric Power University, Jilin, People’s Republic of China.
Email: 2201800147@neepu.edu.cn

Abstract
As an important equipment of power system, the insulator’s normal operation is the basis to ensure the safe operation of the power system. The insulator positioning and identification technology based on machine vision can quickly and accurately complete the inspection of insulators on site and effectively save the cost of operation and maintenance. This paper proposes an insulator inspection method based on region-convolutional neural networks (RCNNs). First, the dataset of the insulator image is preprocessed by means of data expansion. Then, the feature extraction of the insulator image is realised by using the zeiler fergus (ZF) network. The k-means clustering method is used to optimise the selection of anchor points. Meanwhile, the non-maximum suppression post-processing algorithm is improved, and a non-linear penalty factor is introduced to adapt to multi-scale and overlapping occlusion insulator inspection. Experimental results show that the improved faster RCNNs insulator inspection method can accurately obtain the coordinate frame and the corresponding probability value of the insulator object and improve the average precision by 10.43%, achieving the accurate inspection of the insulator object.

1 INTRODUCTION

As an important equipment in power system, insulators are widely used in all links of transmission lines. Insulator equipment is mainly used to support, connect and insulate transmission lines. Insulator deterioration will have an important impact on the safety of the power system. The traditional method for insulator inspection is to use manual inspection in the field. With the rapid development of unmanned aerial vehicle technology, robot technology and computer image processing technology, at present, imaging equipment can be mounted on unmanned aerial vehicle (UAV) and inspection robot to realise automatic inspection of insulators, and thus saving a lot of human costs.

In recent years, object detection and location methods based on the convolutional neural network (CNN) have been widely studied, and great progress has been made in face recognition, speech recognition, traffic sign recognition and other fields. With the maturity and rapid development of artificial neural network technology in image object positioning, the machine vision recognition method for power equipment has also received the attention of scholars. In [1], a detection method of insulators based on local feature and spatial sequence feature is proposed. This method uses multi-scale and multi-feature descriptors to describe the local features of insulators in aerial images, improves the robustness of the algorithm through spatial sequential features, and improves the matching performance of feature points through a coarse precision matching strategy. In [2], an insulator string detection method based on feature point detection and binary local feature detection is proposed. In this method, a double odd-even morphological gradient edge detection algorithm is proposed to complete the image segmentation of insulator strings. Reference [4] proposes an insulator inspection method based on a deep learning method. In this method, a single multi-frame detection (SSD) algorithm is adopted, and the recognition accuracy of porcelain vases and conforming insulators can reach 94.1% and 86.7% respectively. In [5], you only look once version3 (YOLOV3) is mainly used to realise the detection algorithm framework of insulation equipment. Under the condition of high recognition accuracy, the
average detection time of the object is 0.0182 s. In [6], an algorithm is proposed to decompose the object into multiple variable parts with an intersection, and then aggregate them after detection so as to improve the detection accuracy of an aerial insulator. In [7], the insulator is extracted from the detection image taken by UAV for image segmentation. It is used for fault detection of insulators to determine whether the insulator string has self-explosion defects by the distance between the insulators. Reference [8] proposes an automatic positioning method for aerial image insulators based on the binarisation normalised gradient feature and CNN, which is used to detect aerial photo insulators.

In this paper, the algorithm is improved and optimised for the problems such as different sizes, different lengths and widths, overlapping occlusion and image missing in the detection of insulators. The appropriate length-width ratio and the image scale is found by clustering the enhanced datasets, and the post-processing algorithm of non-maximum suppression (NMS) is improved and optimised, which effectively solved the above problems and improved the detection accuracy.

2  |  CORRELATION ALGORITHM OF CNN

2.1  |  Development of object detection methods

Object detection based on traditional manual features needs to diversify detection algorithms and find ingenious computing methods to accelerate the model. Viola–Jones (VJ) algorithm [9] was proposed in IEEE Conference on Computer Vision and Pattern Recognition (CVPR) in 2001, which is a classical face detection algorithm. VJ uses the most conservative sliding window detection, but it is computatively expensive. VJ uses an integral graph to accelerate and overcomplete random Haar features, and adaptive boosting (AdaBoost) algorithm to select features. The histogram of oriented gradient (HOG) [10] in 2005 adopts the most original multi-scale pyramid and sliding window for detection. The size of the detector window is fixed and the multi-scale image pyramid is constructed by scaling the image successively. In order to give consideration to speed and performance, the classifier used by a HOG is usually a linear classifier or a cascade decision classifier. In order to take into account the invariance and discriminative HOG feature, the concept of a block is introduced by a histogram of gradient direction. The 2008 deformable part model (DPM) [11] is based on the peak of classic manual features. DPM splits the detection problem of the whole object in the traditional object detection algorithm and transforms it into the detection problem of each part of the model, and then aggregates the detection results of each part to get the final detection result, which is a whole-part-whole process. There are two main problems in traditional manual feature insulator inspection. First, the region selection strategy based on the sliding window is not objected, the time complexity is high, and the window is redundant. Second, the characteristics of manual design do not have good robustness for the variation of diversity. Especially when dealing with multi-object images, the problem is more prominent.

In 2013, a deep learning-based region-CNNs (R-CNN) series of object proposal object detection was proposed. R-CNN [12] uses a very simple detection strategy. First, the selective search (SS) algorithm is used to generate object proposals in the image based on information such as texture, and then each proposal is scaled to the same size. Second, the network model trained by ImageNet is used for feature extraction. Finally, the support vector machine (SVM) classifier is used to determine the category, the candidate boxes are sorted according to the score, and then remove the excluded candidates boxes according to the rules and return to the last correct box. Although R-CNN has made great progress, its shortcomings are also obvious. Because it is necessary to access the fully connected layer, the size of the picture must be fixed. The preselection box is cropped or stretched and distorted, which will deform the picture and cause distortion. In 2014, the spatial pyramid pooling (SPP) Net [13] added an SPP layer between the convolutional layer and the fully connected layer. The three-level pyramid layers are pooled separately, and the pooled results are stitched and sent to the fully connected layer, but it is still slow. In 2015, Fast R-CNN [14] performed feature extraction on the full graph only and used the SS algorithm to generate the candidate box. The feature of the candidate box is mapped to the feature layer through the position in the original image, and the corresponding area is selected according to the candidate box on the feature map. However, because the candidate box, selected by SS are of different sizes, a fixed-size pooling layer (region of interest (ROI) pooling) is added to extract the fixed-length features (ROI) proposed for each region. At the same time, the multi-task loss function is used, which is connected to the softmax classifier and the box bounding regressor, respectively, to integrate border regression into the R-CNN network. Fast R-CNN adopts the advantages of R-CNN and SPP Net but still needs to use external algorithms to extract candidate boxes. Also in 2015, Faster R-CNN [15] achieved end-to-end processing. From R-CNN to Fast R-CNN to Faster R-CNN, the speed and accuracy of detection have been continuously improved.

In 2015, object detection YOLO of integrated convolution network based on deep learning [16] was proposed. YOLO directly takes the entire image as an input to the network and gets the location of the object bounding box and the category of the object directly through only one forward propagation. Although YOLO has a very fast detection speed, its accuracy is somewhat lower than that of Faster R-CNN, especially the poor detection effect of small objects and its slight insufficiency of positioning accuracy. The SSD [17] algorithm proposed in 2016 has improved these problems by absorbing the advantages of YOLO’s fast speed and accurate region proposal network (RPN) positioning and combining YOLO’s regression idea with Faster R-CNN’s anchor box mechanism. The multi-reference window technique in RPN is adopted, and the detection on a multi-resolution feature map is further proposed. SSD achieved only a lower accuracy on visual object classes 2007 (VOC07) than Faster R-CNN due to its use of multi-level feature classification. As a result, it is difficult to detect small objects. The
Faster R-CNN is composed of RPN and Fast R-CNN. Figure 1 shows the structure of Faster R-CNN. They share a convolutional network layer, which makes the features of small objects not obvious. In 2017, deconvolutional single shot detector (DSSD) [18] changed the reference network of SSD from visual geometry group network (VGG) to resnet-101 in order to solve the problem of poor robustness of small objects, which enhanced the feature extraction ability and added a large amount of context information with deconvolution layer. The prior box selection of multi-dimensional clustering in YOLOv2 [19] plays a better role in improving the aspect ratio. YOLOv3 [20] in CVPR 2018 refers to the residual network structure, forms a deeper network layer, as well as multi-scale detection, and improves the detection effect of mean average precision (AP) and small objects, but its detection of medium or large objects is still unsatisfactory. M2Det [21] in CVPR 2019 introduces a multi-level feature pyramid network on the basis of SSD, which improves the detection of multi-scale objects.

Although the detection model based on an integrated convolutional network is obviously faster than the detection algorithm based on object proposal, its detection accuracy has been slightly inferior to the latter.

### 2.2 Faster R-CNN

Faster R-CNN is composed of RPN and Fast R-CNN. Figure 1 shows the structure of Faster R-CNN. They share a convolutional layer. After the image to be inspected is normalised, it enters a CNN model for feature extraction, and then RPN is responsible for finding the proposal. RPN can greatly reduce the speed of extracting region proposals. After the last shared convolution layer, two branches are used for regression and prediction, respectively, and an ‘anchor (multi-reference window)’ mechanism is designed in it. Fast R-CNN is responsible for further optimising the results of RPN and finally outputting the probability value and coordinates of the object. Among them, the loss function can show the predictive ability of the model and the actual difference, which plays a vital role in the evaluation and adjustment of the model parameters. The loss function consists of two parts, one is the classification loss function, and the other is the regression loss function. From [14, 15], we can get

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

(1)

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

(2)

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

(3)

\[
\text{smooth}_{b_{L1}}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases}
\]

(4)

\[
t_i = (t_x, t_y, t_w, t_h) \quad t_i^* = (t_x^*, t_y^*, t_w^*, t_h^*)
\]

(5)

\[
t_x = (x - x_c)/w_a \quad t_y = (y - y_c)/b_a
\]

\[
t_w = \log(w/w_a) \quad t_h = \log(b/b_a)
\]

(6)

\[
t_x^* = (x^* - x_c)/w_a \quad t_y^* = (y^* - y_c)/b_a
\]

\[
t_w^* = \log(w*/w_a) \quad t_h^* = \log(b*/b_a)
\]

(7)

where \(i\) is an anchor point and \(p_i\) represents the possibility of being predicted as the object. If the anchor is positive, \(p_i^*\) is equal to 1; otherwise it is 0 for classification judgment. \(t_i\) is the coordinate of four parameters \((x_c, y_c, w_a, b_a)\), which represents the coordinate parameters of the network prediction proposal. \(t_i^*\) represents the regression object \((x^*_c, y^*_c, w^*_a, b^*_a)\), and the coordinate parameter of the ground truth box. \(N_{cls}\) and \(N_{reg}\) are used for standardisation. \(L_{cls}\) and \(L_{reg}\) represent classification and regression losses, respectively. \(L_{cls}\) is a logarithmic loss, and \(R\) in \(L_{reg}\) is \(\text{smooth}_{b_{L1}}\) function. The balance parameter \(\lambda\) is used to adjust the weight of the two losses so that they are approximately the same. By default, we set \(\lambda = 10. (x_c, y_c, w_a, b_a)\) represents the coordinate parameter of the anchor box. The difference between the ground truth box and the predicted anchor box was used to study, and the position was regressed and refined. \(R\) is the \(\text{smooth}_{b_{L1}}\) function to reduce the difference between prediction and reality, multiplied by \(p_i\) after the position regression function to represent that it is only activated for the positive anchor. That is, \(p_i^*\) is 1 when there is an insulator, and 0 when there is no insulator, so only the object area calculates the loss, and the unrelated background area does not calculate the loss. Also, \(\lambda\) is set to adjust the weight.
3 | IMPROVED FASTER R-CNN INSULATOR INSPECTION ALGORITHM

3.1 | Dataset enhancement and processing

By transforming the training images, a network with stronger generalisation ability can be obtained, which can better adapt to the application scene, increase the noise data, and improve the robustness of the model. The method in [22] can effectively enhance the dataset. The initial dataset is 500 pictures collected from the training base of our school. This initial dataset is used for data enhancement.

In this paper, the supervised single sample data enhancement technology is adopted. First, a horizontal flip is carried out to increase the diversity of images through mirror operation while increasing the dataset. Then, the sliding window method is used to expand the dataset on a large scale. The sliding window method first normalises the size of all images to 1500*1125 or 1125*1500 pixels. The window with width and height of $W \times H$ is used to slide on each picture, and the region generated by each slide is an ROI region. Equation (8) can be obtained from [22]:

$$\begin{align*}
\text{ROI}_{p_{UL}}X &= I \times W \\
\text{ROI}_{p_{UL}}Y &= J \times H \\
\text{ROI}_{p_{RL}}X &= I \times W + W \\
\text{ROI}_{p_{RL}}Y &= J \times H + H
\end{align*}$$

where $\text{ROI}_{p_{UL}}X$ and $\text{ROI}_{p_{UL}}Y$ are the upper left coordinates of the ROI region. $\text{ROI}_{p_{RL}}X$ and $\text{ROI}_{p_{RL}}Y$ are the coordinates of the lower right corner of the ROI region. $I \in \{0, 1, \ldots, PW/W\}$, $J \in \{0, 1, \ldots, PH/H\}$, $PW$, $PH$ is the width and height of the slide image, $PW = 1500$, $PH = 1125$, or $PW = 1125$, $PH = 1500$; $W = 500$, $H = 375$, or $W = 375$, $H = 500$. Divide the 500 pictures into two parts, one part 400 and the other 100. At the same time, data enhancement was carried out on the two parts of the picture, and 7200 picture sets A and 1800 picture sets B could be obtained, respectively. A is the training set and B is the test set. In order to speed up inspection and training and reduce CPU time consumption, datasets with better practicability for caffe should be produced. All the images in the sample library were manually labelled by the visual image calibration tool software LabelImg to reach the XML label file conforming to the pattern analysis, statistical modelling and computational learning visual object classes (PASCAL VOC) standard format. Through excel, the processed tag file can quickly and intuitively see important information such as image width, height, category name and frame coordinate pairs xmin, ymin, xmax, ymax, and so forth. The completed dataset is processed to fit a better lightning memory-mapped database (LMDB) format.

3.2 | Lightweight network model

The neural network model is the most basic and important part of Faster R-CNN. Generally, the VGG16 network is selected, with a total of 16 convolutional layers and three fully connected layers. However, due to the deep layer of the network, it takes a long time. In addition, the requirements for the graphics processing unit (GPU) are also high, and the network is more suitable for multi-class classification tasks. Considering the GPU configuration of industrial devices in insulator inspection, the singleness of the insulator inspection category, and the fast speed of inspection, it is more appropriate to choose the lightweight network structure model ZF. Figure 2 shows the main network structures including data layers, convolution layers, pooling layers, Relu, fully connected layers, and softmax. Since a larger convolution kernel and stride will lead to unsatisfactory extracted features, convolutional layer uses a 7*7 size convolution kernel with a step size of 2, which helps to improve the classification performance. Then, the activation function Relu is used to increase the non-linearity and sparseness of the network, reduce the interdependence between parameters, and alleviate overfitting. Then, through 3*3 max pooling, translation invariant is used to downsample the image while trying not to lose image features. The advantage is to retain the strongest of these features and reduce the number of parameters. The operation is repeated on the 2, 3, 4, and 5 convolution layers to continuously extract features. After passing through two fully connected layers, local features are integrated through a weight matrix. Finally, it is classified by the softmax function to get the output. Because the network depth is greatly reduced, the inspection speed can be increased, and higher performance in the industry may be achievable.

3.3 | K-means clustering selected anchor

RPN is the core component of Faster R-CNN. The characteristic graph of the output after convolution pooling is convoluted by a 3*3 convolution kernel and then connected to two convolution kernels of 1*1 size, respectively. Among them, a convolution kernel of 1*1 size is a dichotomy with softmax function, and another convolution kernel is a dichotomy with regression
function to determine coordinates. Here, every time the sliding window of size $3 \times 3$ slides, an anchor will be generated, and at the same time, nine anchor boxes with three scales and three proportions will be mapped on the original drawing. As the size and aspect ratio are selected and preset by hand, although the network can learn to adjust the boundary box properly, in practice, there will be inaccurate phenomena in the inspection of some special insulators. If we choose a better prior setting for the network, that is, preselect the specific dataset for a specific purpose, it can make the network easier to learn in the face of complex situations for better prediction and inspection. k-means clustering is run on the training set bounding box to find a good aspect ratio and quantity. According to [19, 20], if the Euclidean distance standard k-means is used, it will cause more errors in the larger boundary box than in the smaller boundary box, and what we want is the case with a higher intersection-over-union (IOU) score:

$$b_1 = (w_1, h_1), b_2 = (w_2, h_2)$$

$$IOU(b_1, b_2) = \frac{\text{area}(b_1 \cap b_2)}{\text{area}(b_1 \cup b_2)}$$

$$d(b_1, b_2) = 1 - IOU(b_1, b_2)$$

Here, $b_1, b_2$ are two boxes, and $IOU$ is the intersection of areas. In this case, the smaller the $d$ is, the higher the $IOU$ is. In Figure 3, k-means is run with different $k$ values and the geometric centre is drawn of the average IOU and the closest. It can be seen that although $k = 5$ gives consideration to both complexity and high recall rate, the recall rate still increases with the increase of dimensions. Therefore, in order to pursue higher inspection accuracy, we further improve $k$. When $k = 15$, the slope approaches 0, indicating that inspection accuracy does not increase with the increase of clustering complexity. So we set $k = 15$. Figure 4 shows the anchor before and after the improvement. The number of anchors has increased from 9 to 15. First, $k$ clusters are randomly assigned, the points divided into the nearest cluster, each cluster updated to the mean value of the current cluster, and iteration continued until the cluster centre is almost unchanged. In the end, we get the aspect ratio as \{0.303, 0.389, 1, 2.57, 3.30\}, and the scale is \{64, 128, 256\}.

3.4 | Non-linear NMS

Due to the anchor mechanism, each anchor will generate a large number of anchor boxes, so the result of multiple anchor boxes of the same insulator will appear in the post-processing stage, that is, repeated inspection. Therefore, NMS processing is required to ensure that an insulator has only one inspection frame.

The greedy NMS is to first select an anchor box $M$ with the highest confidence. Then, the IOU of $M$ and any anchor boxes is traversed, and then all anchor boxes that are greater than the set IOU threshold of $M$ are directly deleted so as to delete all duplicate anchor boxes that are considered to be the same object. However, when multiple objects overlap, that is, when the IOU is greater than the set threshold but not the same object, the anchor boxes are also directly deleted, resulting in a missed inspection. In view of this situation, reference [23] proposed improvements to the NMS method.

In this paper, the non-linear NMS algorithm (NL-NMS) will be selected as the penalty factor according to the IOU and preset threshold of any anchor box $b_i$ and the anchor box $M$ with the highest confidence degree to reduce the confidence $S_i$ of the anchor box. The bigger the IOU between the windows, the heavier the punishment; and the smaller the IOU between the windows, the lighter the punishment. Instead of directly

\[ S_i = S_i \times \left(1 - \frac{IOU(b_i, M)}{IOU_{\text{threshold}}}ight) \]
excluding the boxes that overlap with the checkboxes by more
than a certain threshold, the boxes that are positioned correctly
in crowded situations are not deleted too much. The penalty
factor function is shown as follows:

\[ S_i = S \sqrt{1 - IOU(M, bi)^2} \]  \hspace{1cm} (12)

According to the dynamic change of IOU, the larger the IOU,
the larger the penalty factor value; and the smaller the IOU, the
smaller the penalty factor value. Figure 5 shows the function
curves of various penalty factors, where \( Nl \) is the non-linear
penalty factor in the paper, the Gaussian is a Gaussian curve
with a mean value of 0 and variance of 1, and linear is a lin-
ear curve. It can be seen from the figure that the penalty factor
of the linear curve keeps increasing. The Gaussian curve is still
very high when the IOU is 1. More reasonable is the non-linear
penalty factor in this paper, its inspection effect is better.

4 | EXPERIMENT AND ANALYSIS

4.1 | Experimental process

Faster R-CNN adopted ZF as the training model. In-depth
learning caffe commonly used training set VOC 2007 format
was used as the training dataset format. Table 1 shows the
training parameters, where base_lr represents the base learning
rate; lr_policy represents the adjustment strategy of the learn-
ing rate. Gamma represents the learning rate adjustment factor.
The choice of the above three parameters is based on the empirical
settings in many classic papers, and the experiments in this
article have also confirmed that this choice is appropriate. Step-
size represents the learning rate adjustment stepsize. Its size set-
ing is determined according to the previous pre-training effect.
Batch size stands for batch capacity. A large batch size can reduce
training time and improve stability, but at the same time, it will
also cause the generalisation ability of the model to decrease.
Considering a compromise, we chose 16. Display represents the
parameter display frequency. Setting it to 800 is for us to observe
the loss value as it is unnecessary to wait too long. Iterations
represent the number of iterations. From the loss curve, we can
observe that the curve fully converges when 80,000 iterations
are carried out, so we stop the training. The snapshot represents
the model’s saved node. This is to prevent a sudden break in
training where we do not have to retrain. The test set data is
used to verify the model, and finally, the test model is generated.
The hardware environment built is an Intel processor with a
CPU of 3.0 GHz and 16 GB of memory. GPU for nvidia GTX
1080 graphics processor; the operating system is ubuntu16.04
of Linux, and the model is trained at caffe.

Table 1 Parameter setting table

| The parameter name | The parameter value |
|--------------------|---------------------|
| base_lr            | 0.001               |
| lr_policy          | Step                |
| Gamma              | 0.1                 |
| Step-size          | 20000               |
| Batch size         | 16                  |
| Display            | 800                 |
| Iterations         | 80000               |
| Snapshot           | 10000               |

Figure 6 shows the result of visualisation of convolution
features. From the figure below, we can see the process of
convolution and pooling of feature extraction step by step. The ZF network obtains a visual image after depooling, deelu, and deconvolution on the feature map. The main learning at the bottom is low-level features, convolution 1 mainly learns the physical colour and other characteristics of insulators. Convolution 2 mainly learns the edge and contour characteristics of insulators. Convolution 3 mainly learns the texture characteristics of insulators. From pool1 and pool2, we can clearly see the retained enhanced features. At a high level, the main learning is the abstract features related to insulators. Convolution 4 mainly learns significant distinguishing features. Convolution 5 mainly learns the complete and discriminative key features, which are the characteristics of the entire insulator. In the output, we can see that the basic characteristics of the insulator have been fully displayed. The low-level features can converge when the number of iterations is relatively small, and the high-level requires more iterations. The higher the feature, the better the performance for classification.

4.2 Actual inspection effect

Figure 7 shows that the results were tested using the Faster R-CNN detector before and after the improvement. The confidence level marked in blue on the left of each sub-picture is the inspection effect picture before improvement. The confidence level marked in green on the right is the improved inspection effect picture. At the top of the left picture in (a) and (b), on the right of the left picture in (c), and at the bottom of the left picture in (d), there is a small part of the insulator that has not been detected. Similarly, in the picture on the left of (e), an insulator, half of which is blocked by the equipment, has not been successfully detected. This is because the insulator information is not complete, and the object is too small and missed inspection occurs. In addition, in each of the left images of (f)–(h), there are several insulators that have not been detected correctly. This is because the distance between the insulators is too far, and the object is too small and missed inspection occurs. The improved k-means anchor scheme optimises the selection of anchors according to the ratio and size of the insulators, and at the same time increases the number of anchors, which solves the above problems. There is no false inspection in the right picture in (a)–(h). The three insulators in the left picture in (i) have four inspection frames, and there is a false inspection, which should be caused by the improper setting of the NMS threshold. The insulators on the left in (j)–(l) all have missed inspection. This is a misinspection caused by overlapping insulators. Our improved NI-NMS algorithm can effectively solve the above problems so that there is no false inspection in the right picture of (j)–(l).

4.3 The evaluation index

In this paper, three indexes, including Loss curve, P-R curve and AP, are used to evaluate the inspection effect to verify the improved effect.

4.3.1 Loss curve

Figure 8 shows the Loss curve and learning rate curve of the improved Faster R-CNN for 80,000 iterations, and the Loss curve is generated by parameter output every 800 iterations. As you can see, the curve converges at 66,000 iterations.
4.3.2 P-R curve

P-R is a curve describing the relationship between precision and recall. Precision is defined as the number of true positives over the number of true positives plus the number of false positives. Recall is defined as the number of true positives over the number of true positives plus the number of false negatives. As shown in the following equation:

\[
\text{Precision} = \frac{TP}{TP + FP} \quad (13)
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \quad (14)
\]

Here, \( TP \) is retrieved as a positive example and is also a positive example. \( FP \) is retrieved as a positive example but a negative example. \( FN \) means not retrieved as a positive example, but it is.

Figure 9 shows the P-R curve before and after the algorithm improvement. It can be clearly seen that the area under the curve in the improved figure on the right is larger, which means the inspection effect is better.

4.3.3 AP comparison

AP is a performance metric for predicting object location and category. The larger the AP, the higher the accuracy of inspection. The highest value of AP is 0.9188 after 70,000 and 80,000 iterations. This is because the model has converged after 66,000 iterations, indicating that the inspection effect of the model is getting better and better as the number of iterations increases. The inspection image in Section 4.2 uses the model generated after 80,000 iterations.

The performance of the average IOU at \( k = 5 \) selected by k-means is similar to the preset 9 anchor boxes. We set \( k = 15 \), and Table 2 shows the comparison between Faster R-CNN after using the k-means algorithm to select the scale ratio and AP before the improvement. As can be seen, AP changed from 0.8145 to 0.8776, an increase of 6.31%. This is because the k-means anchor scheme selects and increases the appropriate aspect ratio, which is more suitable for insulator inspection, and detects smaller insulator images that have not been detected previously.

Table 3 shows the comparison between Faster R-CNN of NMS post-processing algorithm with different penalty factors and AP before the improvement, and AP increases from 0.8145 to 0.8903, an increase of 7.58%. This is because the improved NMS algorithm can optimise the occlusion of insulator overlap and reduce missed and false inspection caused by improper threshold selection, thereby improving inspection accuracy.
TABLE 3 Different scales scale anchor boxes’ AP

| Anchor                                | Iterations | AP   |
|---------------------------------------|------------|------|
| Greed non-maximum suppression (NMS)   | 80,000     | 0.8145 |
| Linear NMS                            | 80,000     | 0.8753 |
| Gaussian NMS                          | 80,000     | 0.8845 |
| Non-linear NMS (NI-NMS)               | 80,000     | 0.8903 |

Table 4 lists the comparison of improved k-means anchor and NI-NMS algorithm and AP after the improvement, respectively. AP gradually increased from the original 0.8145 to 0.9188, an increase of 10.43%. This is because we have improved the two algorithms at the same time, so the inspection accuracy has reached the best. At the same time, we can see the same improvement result from the AP curve in Figure 9. In addition, in the actual inspection of Figure 7, the effect of algorithm improvement can also be seen by comparing the inspection of insulators.

5 CONCLUSION

An insulator inspection algorithm based on Faster R-CNN is proposed in this paper. The algorithm first expands the dataset. Then, a lightweight network structure is selected. Meanwhile, k-means algorithm is used to adjust the size scale and ratio of length and width of anchor points to increase the number of object frames of each anchor to 15. The non-linear penalty factor is added to improve the NMS post-processing algorithm. The experimental results after the adjustment and optimisation of the training parameters show that the improved Faster R-CNN has obvious advantages in the field image inspection effect, compared with that before the improvement, and the increase of AP is more than 10%, which can effectively detect the insulator successfully and reduce the error inspection rate and omission rate.

ACKNOWLEDGEMENTS

This work was financially supported by Science and Technology Development Project of Jilin Province (no. 20200403075SF), Jilin Provincial Department of Science and Technology (no. 20180201010GX), Jilin Provincial Department of Education (no. JJKH20180440KJ).

REFERENCES

1. Liao, S.; An, J.: A robust insulator detection algorithm based on local features and spatial orders for aerial images. IEEE Geosci. Remote Sens. Lett. 12(5), 963–967 (2015)
2. Zhao, Z., Xu, G., Qi, Y.: Representation of binary feature pooling for detection of insulator strings in infrared images. IEEE Trans. Dielectr. Electr. Insul. 23(5), 2858–2866 (2016)
3. Yin, J., et al.: Edge detection of high-voltage porcelain insulators in infrared image using dual parity morphological gradients. IEEE Access 7, 32728–32734 (2019)
4. Miao, X., et al.: Insulator detection in aerial images for transmission line inspection using single shot multibox detector. IEEE Access 7, 9945–9956 (2019)
5. Liu, Y., et al.: Research on automatic location and recognition of insulators in substation based on YOLOv3. High Voltage 5, 62–68 (2020)
6. Gao, Q., Liu, Q.: Research on object detection of insulator in aerial image. Electr. Meas. Instrum. 56(05), 119–123 (2019)
7. Guo, T., et al.: Self-explosion defect identification of insulator based on faster RCNN. Insul. Surge Arresters 3, 183–189 (2019)
8. Peng, X., Liu, Y., Wang, K.: An automatically locating method for insulator object based on CNNs. Geomatics Inf. Sci. Wuhan Univ. 44(04), 563–569 (2019)
9. Viola, P.A., Jones, M.J.: Rapid object detection using a boosted cascade of simple features. In: Computer Vision and Pattern Recognition, Kauai, HI, USA (2001)
10. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision & Pattern Recognition, San Diego, CA, USA (2005)
11. Felzenszwalb, P., Mcallester, D., Ramanan, D.: A discriminatively trained, multiscale, deformable part model. In: 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA (2008)
12. Girshick, R., Donahue, J., Darrell, T.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA (2014)
13. He, K., Zhang, X., Ren, S.: Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. 37(09), 1904–1916 (2015)
14. Girshick, R.: Fast R-CNN. In: 2015 IEEE International Conference on Computer Vision, Santiago, Chile (2015)
15. Ren, S., et al.: Faster R-CNN: Towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, Montreal, Canada, (2015)
16. Redmon, J., Farhadi, A., Girshick, R.: You only look once: Unified, Real-time object detection. In: IEEE Conference on Computer Vision & Pattern Recognition, Boston, MA, USA (2015)
17. Liu, W., et al.: SSD: Single shot multibox detector. In: European Conference on Computer Vision, Amsterdam, Holland (2016)
18. Fu, Y., et al.: DSSD: Deconvolutional single shot detector. In: 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA (2008)
19. Redmon, J., Farhadi, A.: YOLO9000: Better, faster, stronger. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA (2017)
20. Redmon, J., Farhadi, A.: YOLOv3: An incremental improvement. In: IEEE International Conference on Computer Vision, Seoul, South Korea (2019)
21. Zhao, Q., et al.: M2Det: A single-shot object detector based on multi-level feature pyramid network. In: The Association for the Advance of Artificial Intelligence, Honolulu, HI, USA (2019)
22. Li, L., Ma, W., Li, L.: Research on detection algorithm for bridge cracks based on deep learning. Acta Autom. Sin. 45(09), 1727–1742 (2019)

23. Bodla, N., Singh, B., Chellappa, R.: Soft-NMS—improving object detection with one line of code. In: 2017 IEEE International Conference on Computer Vision, Venice, Italy (2017)

How to cite this article: Wu J, Tang S, Li X. An insulator inspection method based on deep learning applicable to multi-scale and occlusion conditions. J Eng. 2021;2021:216–225.
https://doi.org/10.1049/tje2.12029