Detection Method of Insulator Based on Single Shot MultiBox Detector

Changfu Xu, Bin Bo, Yang Liu and Fengbo Tao

State Grid Jiangsu Electric Power Co., Ltd. Research Institute, Nanjing 211103, China
xuchangfu2008@126.com

Abstract. Insulators are the most common equipment in the power system, the failure of insulators will cause heavy economic loss to electric power companies, so it is very important to detect insulators effectively for inspecting their working states. This paper proposes a novel method to detect the insulators based on single shot multibox detector (SSD) in which discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. A large number of visible light images are used as experimental data in experiment, and the results show that this method can detect small-size insulators in complex background with high precision as well as low time cost.

1. Introduction

With the development of intelligent grid and power system automation, computer vision techniques are increasingly applied on the intelligent inspection and online monitoring for equipment on power system. Insulators are indispensable insulation components in power transmission lines, and its operating conditions directly affect the reliability and safety of power grids. At the same time, insulators play the role of electrical insulation and supporting in the transmission lines [1-2]. And the contamination, cracks, and aging on the surface of the insulators seriously threaten the safe operation of transmission lines. According to statistics, the highest percentage of accidents in the current power system failure is caused by insulator defects. Therefore, it is very important to monitor the condition of the insulator.

Some methods of detecting the insulators have been proposed. Zhang [3] describes a method based on Harris corner matching and spectral clustering. Harris corner detection is performed on the selected template image and the image to be detected, respectively; the fuzzy clustering is conducted on the matched corners, the cluster centers are extracted, and the insulator contour is extracted with Chan_Vese model. Wu proposes a texture segmentation algorithm based on active contour to divide the insulator images into sub-regions. In their experiment, they used the insulator images in which the texture features of insulators are very different from those of the background, and active contours needs to be set and this influences the real-time performance[4]. REDDY [5] uses K-means clustering algorithm to segment the insulator image into different classes, then annotates each candidate area image block which is sent to the support vector machine for classification. This method can detect insulators precisely when the background is simple but it performs badly when the background is complex. YUAN [6] uses ASIIFT algorithm to realize the detection of insulators. A standard insulator gallery is established. Then the transmission line video and the images in the established gallery are matched by ASIIFT that the identification and location of insulators can be done. Zhao [7] presents an approach for extracting the edge of insulators based on NSCT (non subsampled contourlet transform). This method uses NCST to split image into blocks, the local threshold value of every coefficient block is calculated, and the binary edge image is obtained based on the threshold. Although the above
method can identify the insulator in the image, but due to the background of the image is complicated and it includes pseudo-targets such as towers and lines. This characteristic make above algorithm less accurate. And their computation and complexity increase greatly when the datasets grow larger.

In recent years, with the increase of computer computing capabilities, deep learning [8] have received more and more attention from researchers. The deep learning methods such as Convolution Neural Networks have an increasing application in image classification, speech recognition, and target detection. These applications also allow us to see the bright future of deep learning in power technology. Inspired by deep learning, we apply SSD algorithm [9] to recognize and locate insulators in images with complicated background. This method discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. It achieves good detection accuracy and has advantages of low computational and low time consuming for the detection of insulators with complicated background. It also can lay the foundation of judging the working states of insulators.

The rest of this paper is organized as follows. Section 2 introduces the proposed method applied to detect insulators. We analyze experiment results from precision and speed in Section 3 and finally conclude this work.

2. Detection Method

Fig. 1 shows the architecture of automatic insulator detection method based on SSD. This method is based on feed-forward convolutional neural network. It makes the detection procedure in one network.

Figure 1. The algorithm architecture of automatic detection of insulators.

The early network layers are based on a standard architecture VGG16 which is used for high quality image classification. Then we add auxiliary structure to the network to achieve the insulator detection. We add convolutional feature layers to the end of the truncated base network (VGG16). The convolutional layers are used to instead the fully connected layer of VGG16, and then add four convolutional layers.

Table 1 is the structure of the hole SSD netwok. The six different convolution layer feature maps of conv4_3, fc7, conv6_2, conv7_2, conv8_2 and conv9_2 are used for detection. Each feature layer can produce a fixed set of detection predictions by using a set of convolutional filters. These are indicated on top of the SSD network architecture in Fig. 1. For a feature layer that its size is \( m \times n \) with \( p \) channels, the basic element for predicting parameters of a potential detection is a \( 3 \times 3 \times p \) small kernel that produces either a score for a category, or a shape offset relative to the default box coordinates. The default box is a series of fixed-size boxes on each cell of the feature map. The number of default boxes at each grid is \( k \). Each default box has two classes (it is insulator or not) and four offsets. The feature map whose size is \( m \times n \) that has \( m \times n \) feature map meshes. The number of the feature map output is \( 6 \times k \times m \times n \).
In the training stage, we need to choose which default box should match with the ground truth. For each ground truth box we select from default boxes which vary over location, aspect ratio, and scale. Once matched, the default box is a positive sample, if it is not matched; the default box is a negative sample. The negative samples are ranked according to the value of confidence loss. We use the front negative samples which can keep the proportion of positive and negative samples at 3:1. In the test stage, the offset of the default box and the corresponding confidence in the target category are got.

Table 1. Parameters of SSD Network

| Conv layer | kernel | number of kernels | stride | padding | output     |
|------------|--------|------------------|--------|---------|------------|
| Conv1_1    | 3*3    | 64               | 1      | 1       | 300*300    |
| Conv1_2    | 3*3    | 64               | 1      | 1       | 300*300    |
| Maxpool1   | 2*2    | 1                | 2      | 0       | 150*150    |
| Conv2_1    | 3*3    | 128              | 1      | 1       | 150*150    |
| Conv2_2    | 3*3    | 128              | 1      | 1       | 150*150    |
| Maxpool2   | 2*2    | 1                | 2      | 0       | 75*75      |
| Conv3_1    | 3*3    | 256              | 1      | 1       | 75*75      |
| Conv3_2    | 3*3    | 256              | 1      | 1       | 75*75      |
| Conv3_3    | 3*3    | 256              | 1      | 1       | 75*75      |
| Maxpool3   | 2*2    | 1                | 2      | 0       | 38*38      |
| Conv4_1    | 3*3    | 512              | 1      | 1       | 38*38      |
| Conv4_3    | 3*3    | 512              | 1      | 1       | 38*38      |
| Conv4_3    | 3*3    | 512              | 1      | 1       | 38*38      |
| Maxpool4   | 2*2    | 1                | 2      | 0       | 19*19      |
| Conv5_1    | 3*3    | 512              | 1      | 1       | 19*19      |
| Conv5_2    | 3*3    | 512              | 1      | 1       | 19*19      |
| Conv5_3    | 3*3    | 512              | 1      | 1       | 19*19      |
| Maxpool5   | 3*3    | 1                | 1      | 1       | 19*19      |
| Fc6        | 1*1    | 1024             | 1      | 1       | 19*19      |
| Fc7        | 1*1    | 1024             | 1      | 0       | 19*19      |
| Conv6_1    | 1*1    | 256              | 1      | 0       | 19*19      |
| Conv6_2    | 3*3    | 512              | 2      | 1       | 19*19      |
| Conv7_1    | 1*1    | 128              | 1      | 0       | 10*10      |
| Conv7_2    | 3*3    | 256              | 2      | 1       | 10*10      |
| Conv8_1    | 1*1    | 128              | 1      | 0       | 5*5        |
| Conv8_2    | 3*3    | 256              | 1      | 0       | 3*3        |
| Conv9_1    | 1*1    | 128              | 1      | 0       | 1*1        |
| Conv9_2    | 3*3    | 256              | 1      | 0       | 1*1        |

In training the SSD network, the loss function is a weighted sum of the localization loss (loc) and the confidence loss (conf):

$$L(x,c,l,g) = \frac{1}{N} \left( L_{conf}(x,c) + \alpha L_{loc}(x,l,g) \right)$$

where $N$ is the number of matched default boxes. If $N = 0$, then the loss is 0, too. The localization loss is a Smooth L1 loss between the groundtruth box ($g$) and the predicted box ($l$) parameters. Similar to Faster R-CNN [2], offsets are regressed for the center ($cx$, $cy$) of the default bounding box ($d$) and for its width ($w$) and height ($h$).

$$L_{loc}(x,l,g) = \sum_{i \in \text{Pos, m} \in \{cx, cy, w, h\}} \sum_{m} x_g^m \cdot \text{smooth}_{L1} \left( l^m_i - g^m_i \right)$$
\[ g_j^{cx} = \left( g_j^{cx} - d_i^{cx} \right) / d_i^{wx} \]
\[ g_j^{cy} = \left( g_j^{cy} - d_i^{cy} \right) / d_i^{wy} \]
\[ g_j^w = \log \left( \frac{g_j^w}{d_i^{wx}} \right) \]
\[ g_j^h = \log \left( \frac{g_j^h}{d_i^{wy}} \right) \]

The confidence loss is the softmax loss over multiple classes confidences (c).

\[ L_{conf}(x, c) = -\sum_{i \in \text{Pos}} x_i^p \log \left( \sum_{j} c_j^p \right) \]

and the weight term \(\alpha\) is set to 1 by cross validation.

3. Experiment and Analyses

In this section, we conduct experiment on a computer with GTX980ti using our insulator datasets.

3.1. Datasets

Since there are no insulator datasets available from public resources, the experimental datasets we used in this paper are provided by State Grid Jiangsu Electric Power Company Research Institute, but only about 2000 insulator images with complicated background can be used, in which the numbers of insulators are different for each insulator image. Then we use the means of rotating and overturning to make the number of insulator images to 6000. We select randomly 4500 images for training, 800 images for validation and 700 images for test. The graphical image annotation tool LabelImg is used to annotate the insulators in each image and each insulator is surrounded by a rectangular that called ground-truth box. Fig. 2 is the samples of annotated insulators.

3.2. Performance Evaluation Index

In this article, we use the missing detection rate and accuracy to measure the detection algorithm. By testing the model on the insulator datasets, we record each detection box in insulator image and the IoU between detection box and ground truth box.

We assume detection box is BBdt, ground truth box is BBgt. If the IOU value is greater than the threshold, then BBdt and BBgt are matching. We set the IoU as 0.5 in this article. It is defined as follows:

\[ \text{IoU} = \frac{\text{area}(BB_{dt} \cap BB_{gt})}{\text{area}(BB_{dt} \cup BB_{gt})} > 0.5 \]

In the process of BBdt matching with BBgt, the unmatched BBdt is wrong detection insulator (False Positive, FP), the unmatched BBgt is the undetection insulator (False Negative, FN). The standard insulator statistics is shown in Table 2.
The annotated insulator images

**Fig. 2.** The annotated insulator images

| Detection result          | Insulator (Positive) | Not Insulator (Negative) |
|---------------------------|-----------------------|--------------------------|
| **Insulator (Positive)**  | True Positive (TP)    | False Positive (FP)      |
| **Not Insulator (Negative)** | False Negative (FN) | True Negative (TN)       |

The undetection insulator rate is defined as follows:

\[
R = \frac{FN}{FN + TP}
\]
The precision is defined as follows:

\[ P = \frac{TP}{TP + FN} \]

where TP, FN, FP represent the number of True Positive, False Negative and False Positive.

3.3. Training Details

Figure 2 shows the architecture details of the SSD300 model. The conv4_3, conv7 (fc7), conv8_2, conv9_2, conv10_2, and conv11_2 layers are used to predict both insulator confidences and location. We set the scale of default box as 0.1 on conv4_3 layer. We initialize the parameters for all the newly added convolutional layers by using the “xavier” method. For the layer of conv4_3, conv10_2 and conv11_2, we only associate 4 default boxes at each feature map location – omitting aspect ratios of 0.33 and 3. But for all other layers, we put 6 default boxes. Since, the layer of conv4_3 has a different feature scale compared to the other layers, the L2 normalization technique introduced in [10] is used to scale the feature norm at each location in the feature map to 20 and learn the scale by using back propagation. We set the learning rate as 0.001 for 40k iterations, and then continue training for 10k iterations with the learning rate as 0.0001 and 0.00001. The moment is set as 0.9 the weight decay is set as 0.0005.

3.4. Model Comparision

In this paper, we train 2 models on this insulator dataset. As shown in Table 3, the size of the input image will affect the accuracy and the character of real time. The size of the input image in model SSD300*300 is smaller than the model of SSD512*512. Although the average test time is faster than the model of SSD 512*512, but the average test precision is less than the model of SSD 512*512. From Table 3, it is clear that this method can meet the real-time requirements.

| Model       | Detection time(s) | Miss rate (%) |
|-------------|-------------------|---------------|
| SSD 300*300 | 0.03              | 12.1          |
| SSD 512*512 | 0.07              | 10.7          |

3.5. Detection Results and Analyses

We use 500 images to test the performance of our proposed method. The final detection results of our method indicate that it can precisely recognize and locate insulators in an image. Fig 3 illustrates the detection results of four images, and we can see that most of insulators can be detected precisely.
To fully demonstrate the accuracy and speed of our method, we compare our method with other detection methods based on HOG and SVM, ACF, R-CNN, Faster R-CNN. The summarized detection results are listed in Table 4 and Fig4. From Table 4, it is clear that our method outperforms all the other deep learning target detection methods by mean average precision. The method of R-CCN is based on selective search and convolutional neural network [11]. Selective search is based on low-level image feature to generate region proposals. Faster R-CNN [12] replaces selective search proposals by ones learned from a region proposal network (RPN), and introduces a method to integrate the RPN with Fast R-CNN by alternating between finetuning shared convolutional layers and prediction layers for these two networks. This way region proposals are used to pool mid-level features and the final classification step is less expensive.

Our SSD is very similar to the region proposal network (RPN) in Faster R-CNN in that we also use a fixed set of (default) box for prediction, similar to the anchor boxes in the RPN. But instead of using these to pool features and evaluate another classifier, we simultaneously produce a score for each object category in each box. Thus, our approach avoids the complication of merging RPN with Fast R-CNN and is easier to train, faster, and straightforward to integrate in other tasks. Our method achieves a recognition rate with a 5.8% mAP higher than the baseline method using R-CNN. From Table 3, it also can be found that our method takes a mean time of 0.03s per image used for detecting the insulators, while Faster R-CNN takes 0.11s per image when detecting. In terms of detection time, our method shows the importance of using multi-scale convolutional bounding box outputs attached to multiple feature maps at the top of the network. By comparing precisions and recalls in Fig 4, the results illustrate that our method also performs better than the methods based on HOG and SVM, ACF, respectively.

Table 4. The comparison of 3 detection methods

| Methods      | Detection time(s) | Detection speedup | mAP (%) |
|--------------|-------------------|-------------------|---------|
| R-CNN        | 26.56             | 1×                | 86.8    |
| Faster R-CNN | 0.11              | 93.9×             | 92.6    |
| Ours         | 0.03              | 202.8×            | 94.7    |
4. Conclusions
This paper propose an object detection method based on SSD, a fast single-shot object detector for detecting insulators in an image. A key feature of our model is the use of multi-scale convolutional bounding box outputs attached to multiple feature maps at the top of the network. This representation allows us to efficiently model the space of possible box shapes. Experiment on a large number of insulator images show that this method has high detection accuracy and low time cost in recognizing and locating insulators with complex background. In the further study, increasing the diversity and capacity of our datasets and improving the detection accuracy are considered.

5. Acknowledgments
This work is supported by Science and Technology Guide Program of State Grid (SGTYHT/14-JS-191).

6. References
[1] B. Li, D. Wu, Y. Cong, and Y. Xia, “A method of insulator detection from video sequence.” vol. 8330, no. 1, pp. 386-389, 2012.
[2] H. Ha, and J. Lee, “Bad Insulator Detection Algorithm Using Microphone Array,” IEEE/ASME International Conference on Advanced Intelligent Mechatronics, vol. 23, no. 3, pp. 266 - 271, Vol.23, pp. 266 - 271. IEEE.
[3] G. N. Zhang, Z. G. Liu. “Fault detection of catenary insulator damage / foreign material based on corner matching and spectral clustering.” vol. 35, no. 6, pp. 1370 - 1377, 2014.

[4] Q. Wu, J. An, and B. Lin, “A Texture Segmentation Algorithm Based on PCA and Global Minimization Active Contour Model for Aerial Insulator Images,” IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing, vol. 5, no. 5, pp. 1509-1518, 2012.

[5] M. J. B. Reddy, B. K. Chandra, and D. K. Mohanta. “A DOST Based Approach for the Condition Monitoring of 11 kV Distribution Line Insulators,” Dielectrics & Electrical Insulation IEEE Transactions on, vol. 18, no. 2, pp. 588 - 595, 2011.

[6] J. S. Yuan, K. B. Cui, B. S. Li. “Identification and location of insulator video images based on ASIFT algorithm.” Electrical Measurement & Instrumentation. vol. 52, no. 7, pp. 106 - 112, 2015.

[7] Z. Zhao, and N. Liu, “The Recognition and Localization of Insulators Adopting Surf and Ifs Based on Correlation Coefficient,” Optik - International Journal for Light and Electron Optics, vol. 125, no. 20, pp. 6049-6052, 2014.

[8] Lecun Y, Bengio Y, Hinton G. Deep learning [J]. Nature, 2015, 521(7553): 436.

[9] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector [J]. 2015:21-37.

[10] Liu W, Rabinovich A, Berg A C. ParseNet: Looking Wider to See Better [J]. Computer Science, 2015.

[11] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2014:580-587.

[12] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real Time Object Detection with Region Proposal Networks,” IEEE Transactions on Pattern Analysis & Machine Intelligence, 1-1, 2016.