Semantic Segmentation and Defect Detection of Aerial Insulators of Transmission Lines

Shengli Wang¹, Zhangpeng Zhou¹ and Wenbin Zhao²*

¹Maintenance company of State Grid Power Company in Gansu Province, 730000, Lanzhou Gansu, China
²Shanghai University of Electric Power, Shanghai, 200090, China
*Email: wenbinmax@163.com

Abstract. Aiming at the problems of low accuracy and poor generalization ability of insulator defect detection in complex aerial images by existing insulator defect detection algorithms, the possibility of using semantic segmentation technology to simplify insulator features in complex images is explored. The semantic segmentation model DeepLabv3 is cascaded with the target detector yolov3 to realize the semantic segmentation of insulators in aerial images and the detection of defects. The experimental results show that the use of the strategy of semantic segmentation and target detection can increase the accuracy of insulator defect detection by 12.58%, which effectively improves the performance of the detection model.

1. Introduction

In recent years, the scale of my country's power supply has continued to expand, and power technology has also been continuously developed to ensure the stable and effective operation of the power system. Transmission lines are in the natural environment for a long time and are greatly affected by extreme weather and natural disasters, which are likely to cause damage to power devices and cause safety hazards. In order to ensure the safe operation of the power system, it is very important to conduct regular power inspections. In the inspection work, the defect detection of insulators is the most important thing. Insulators are a key part of power transmission lines, and they are used in huge quantities and in various types. On the one hand, it plays the role of mechanical support for the wire, and on the other hand it plays the role of insulation, preventing the current from forming a channel to the ground [1]. Insulators are easily damaged in the field environment, causing mechanical and electrical failures, leading to failure of various electromechanical stresses, thereby damaging the operating life of the entire line, and in severe cases, it may cause the power grid to paralyze. Therefore, timely and accurate detection of insulator defects is particularly important.

The traditional methods of insulator defect detection range from the initial direct observation method to the manual analysis method based on images and videos. Although it is possible to remotely observe whether the insulator has defects, but due to the huge amount of image and video data, manual analysis is time-consuming and laborious [2]. The rise of image recognition technology has accelerated the development of insulator defect detection technology, greatly reduced labor costs, and achieved real-time high-precision detection of insulators on transmission lines. Image processing-based insulator defect detection mostly uses semantic segmentation technology, which can adapt to image data in different environments by using the characteristics of its segmented pixels. Chen et al. [3] proposed a U-net network-based aerial insulator detection method, which extracts features from aerial images hierarchically, and fuses shallow features with deep features. Among them,
shallow features are used to locate pixels, and deep features are used to classify pixels. This method can avoid the loss of detailed information such as the target position, and at the same time improve the accuracy, and can make effective semantic segmentation of insulators in a more complex environment. Li et al. [4] used SSD (Single Shot Multi Box Detector) convolutional neural network method to detect insulator defects in transmission line images. This method uses ResNet to replace the original VGGNet in the SSD network structure as the backbone network, which enhances the feature extraction capability of the network. Finally, the insulator and defect positions detected by the network model are overlapped and calculated to achieve positioning. Wang et al. [5] proposed an insulator defect detection method that combines the shape, color and texture of the insulator. This method perceives parallel line features from different directions in the detection image, and uses it as an insulator candidate area, and then expands the candidate area to adjacent areas, analyzes and recognizes the saliency of the local neighborhood, and uses PCA (Principal Component Analysis) to compensate. The insulator area is then divided into multiple blocks according to the average distance between the insulators, and the defect detection of the insulators is realized by analyzing the texture changes. However, the above-mentioned traditional image processing methods rely more on specific features in the image. When the image processing effect is not good, the applicability of the detection method will also be reduced. Due to its advantages in calculation and feature extraction, image processing based on deep learning has made the image semantic segmentation method based on deep learning a trend in current insulator defect detection research.

Figure 1. Semantic segmentation and defect detection cascade model structure diagram.

Aiming at the specific task of detecting insulator defects in aerial images, this paper selects DeepLabv3 semantic segmentation technology, which has excellent performance in the field of semantic segmentation, to extract insulator masks in complex images, and cascade it with mainstream target detector yolov3, in the target detector features in the extraction stage, the mask data is used to replace the original data to improve the final detection performance of aerial insulator defects.

2. Algorithm
In recent years, semantic segmentation algorithms based on deep learning have mainly relied on the development of deep neural networks, from the early LeNet [6] handwritten digit recognition application to the ImageNet competition champion algorithm AlexNet [7] and VGG [8], and then to the later ResNet [9], FCN [10] (Fully Convolution Network) and other networks are proposed, and the network structure is becoming more and more complex. With the active development of deep learning, semantic segmentation networks based on deep learning are constantly being proposed, such as U-Net, SegNet, DeepLab, PSPNet and other models are widely used. Most of the semantic segmentation networks and models used in electric power inspection aerial images are subjected to experiments and
applications after migration learning or network modification by general semantic segmentation models. Therefore, there are many challenges and breakthroughs in the semantic segmentation technology of electric power inspection aerial images. Including real-time model, model small target segmentation, model multi-scale and contextual information fusion, etc. Therefore, in response to the above problems, this paper extracts the insulator mask features based on the DeepLabv3 semantic segmentation model, compares it with other mainstream models U-Net [11], SegNet [12], PSPNet [13], and combines the target detection algorithm yolov3 to achieve insulator defects Detect and use the mask data to further improve the detection accuracy.

The cascade model structure of insulator semantic segmentation and defect detection is shown in figure 1. In the semantic segmentation process, the performance of different semantic segmentation models is tested. By learning context information, the mask information containing insulator semantics is extracted from multi-scale, and cross-entropy loss is used to optimize the semantic segmentation area. The mask data is further sent to the target detector for defect detection. The mask data is extracted from the backbone network DarkNet-53 to extract the features c1, c2, c3, c4, and c5, and then use the feature pyramid to establish the horizontal connection between the feature maps, which will be integrated. The subsequent feature map output is mapped to Feature Pyramid to get p3, p4, and p5. Finally, the detection head of yolov3 is used to make the final prediction. For classification and positioning, cross-entropy loss and L1 loss are used respectively. The objectiveness loss is used to supervise the confidence score and judge whether there is an object in the target frame.

The DeepLab [14] deep convolutional network model uses hole convolution instead of deconvolution to increase the receptive field and obtain more contextual information. This method also adds Conditional Random Field (CRF) to improve the accuracy of network semantic segmentation. DeepLabv2 [15] proposed the ASPP (Atrous Spatial Pyramid Pooling) module, which uses the expanded convolution kernel with 4 sampling rates to extract multi-scale features, and then uses the fully connected random field to optimize the segmentation effect. DeepLabv3 [16] improves ASPP on this basis, adds a batch normalization layer after the hole convolution, passes the result through 1x1 convolution, and then uses bilinear interpolation to upsampling to obtain the required spatial dimension, segmentation The performance has been significantly improved, and the specific structure is shown in figure 2. The semantic segmentation model uses a combination of codecs, uses an end-to-end training method, and uses the backbone network Xception and ASPP modules as the encoding end to extract features. Among them, Xception is a deep convolutional neural network that uses deep separable convolution to reduce the amount of model parameters.

![Semantic segmentation model structure diagram.](image)

ASPP is a multi-scale pyramid feature extraction module composed of hole convolutions with multiple expansion rates.

The encoding end obtains the input feature layer through the backbone network, and uses the hole convolution with different expansion rates to perform feature extraction on the part with higher
semantic information, and performs 5 parallel operations, respectively 1x1 convolution, and the expansion rate is 6, 12, 18 are 3x3 hole convolution and global average pooling, and then stack the results obtained after these five operations, use a 1x1 convolution to adjust the number of channels, and the part of enhancing feature extraction ends. The part with shallow semantic information is also used to adjust the number of channels by 1x1 convolution, and then merged with the part obtained from enhanced feature extraction to realize the fusion of low-level feature information and high-level feature information.

The decoding end uses a simpler combined up-sampling module to predict the output segmentation results. After integrating the low-level features directly output by Xception and the high-level features output by the encoding end, it performs bilinear interpolation up-sampling, which can get better with less training set The training segmentation effect.

The model uses binary cross entropy loss to calculate the cross entropy between each pixel, the calculation method is as equation (1).

$$\text{Loss} = -\frac{1}{N} \sum_{i=1}^{N} \left[ g_i \times \ln(p_i) + (1 - g_i) \times \ln(1 - p_i) \right]$$

(1)

where \( g_i \) is the true category of pixel \( i \), \( p_i \) is the prediction category of the pixel \( i \) by the model.

3. Experiment and Result Analysis

This article uses the PyTorch open source framework, the test environment is NVIDIA Tesla T4 GPU, the operating system is linux, CUDA 10.1 version, and the compiled language is Python 3.7. A total of 600 aerial images of insulators and 100 images of defective insulators were used in the experiment, and all the images were labeled as examples. Randomly select 500 of them as the training set of the network, and the remaining 100 as the test set. The training process selects the Adam optimizer with faster convergence speed. The initial parameters are set as learning rate, number of iterations, learning rate attenuation factor, and the training is final. To evaluate the effectiveness of the experiments in this paper, root mean square error (RMSE) and mean intersection ratio (MIoU) are used to evaluate the effect of semantic segmentation, as in equation (2-4).

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_{\text{obs},i} - X_{\text{model},i})^2}$$

(2)

$$\text{IoU} = \frac{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ui}}{\sum_{j=0}^{k} p_{ij}}$$

(3)

$$\text{MIoU} = \frac{1}{k + 1} \sum_{j=0}^{k} \text{IoU}_j$$

(4)

where \( k \) is the number of semantic categories, \( i \) is the true value, \( j \) is the predicted value, \( p_{ui} \) is the Real case, \( p_{ij} \) is the False positives, \( p_{ji} \) is the False negatives.

After semantic segmentation obtains the insulator mask map, mark the defective mask map, send it to the target detector for experiment, and use precision and recall to measure the results of the detection model. The algorithm formula is as equation (5-6).

$$\text{precision} = \frac{TP}{TP + FP}$$

(5)

$$\text{recall} = \frac{TP}{TP + FN}$$

(6)

Where TP is positive samples, FP is positive samples with negative values, FN is negative samples with negative values.
Different semantic segmentation models can identify pixels belonging to insulators in aerial images, and the segmentation effect is shown in figure 3.

**Figure 3.** Segmentation renderings of different semantic segmentation models.

Using different semantic segmentation models to segment the test set composed of 100 random images, the experimental results obtained are shown in table 1. Among them, DeepLabv3 has the best segmentation effect, and the RMSE is 0.17 and the MioU is 0.814 better than the other three methods. It increases the receptive field, optimizes the semantic segmentation of small targets, and has a certain context integration capability, which can completely segment a single insulator part in the insulator string, and the rich details are conducive to the subsequent defect detection work. Due to the simple structure of the U-Net network, the under-fitting phenomenon cannot be completely segmented, and a large number of discontinuous pixels appear. High RMSE and low MioU indicate poor segmentation performance. SegNet and PSPNet perform fairly well, relying on the codec structure to effectively segment the insulator body, but the resulting mask lacks details.

**Table 1.** Comparison of semantic segmentation results of different models.

| Model       | RMSE | MioU  |
|-------------|------|-------|
| U-Net       | 0.25 | 0.627 |
| PSPNet      | 0.19 | 0.787 |
| SegNet      | 0.21 | 0.735 |
| DeepLabv3   | 0.17 | 0.814 |

The original data and the corresponding mask data marked with the defect location are input into the target detector for defect detection. The detection results are shown in table 2. The comparison of the defects detected by DeepLabv3 before and after yolov3 is shown in figure 4.

**Table 2.** Comparison of defect detection results under different strategies.

| Strategy | yolov3 | U-Net +yolov3 | PSPNet +yolov3 | SegNet +yolov3 | DeepLabv3 +yolov3 |
|----------|--------|---------------|----------------|----------------|-------------------|
| recall/% | 76.25  | 76.75         | 89.25          | 83.00          | 95.25             |
| precision/% | 86.67 | 85.00        | 94.00          | 88.75          | 99.25             |
Figure 4. Comparison of defect detection results before and after semantic segmentation.

From Table 2 we can see the effect of different strategies on the defect detection effect. When using the mask data obtained from semantic segmentation for training, the features learned by the model will not be affected by the complex background. Compared with the training using raw data, the recall rate is increased by up to 19%, and the accuracy is increased by up to 12.58%. Moreover, different segmentation effects have different effects on the performance of the target detector. The mask image obtained by U-Net has a poor segmentation effect on the insulator string. Using its mask data for training will not improve the accuracy of target detection. The use of insulator mask data obtained by PSPNet, SegNet, and DeepLabv3 all improve defect detection performance to varying degrees. The richer the mask image details and the clearer the edges, the more concise the target features, which can effectively improve the detection accuracy.

4. Conclusion

In order to improve the detection ability of insulator defects in aerial complex images, this paper explores the possibility of semantic segmentation of aerial insulator images, introduces DeepLabv3 semantic segmentation technology in detail, and conducts comparative experiments with mainstream semantic segmentation networks such as U-Net, PSPNet, and SegNet. Analysis and use the above network to obtain the insulator mask map, thereby simplifying the insulator features in the aerial complex images, and effectively improving the detection performance of the yolov3 target detection model for insulator defects. For the semantic segmentation algorithm of aerial insulator images, further research will be done on the lightweight of the model and the real-time and accuracy direction of image processing in the future.

References

[1] Li W X, Zheng W L, Wang C S and et al. 2020 Research on detection technology of suspension points of transmission lines *Instrumentation Users* **27** 16-20
[2] Huang X, Zhao J, Gao F and et al. 2017 Ultrahigh Insulator Detection Robot System Design *DEStech Transactions on Computer Science and Engineering*
[3] Chen J W, Zhou X and Zhang R and et al. 2018 Aerial insulator detection based on U-net network *Journal of Shaanxi University of Science and Technology* **036**(004): 153-157
[4] Li W X, Zheng W L, Wang N and et al. 2019 Research on Defect Detection Method of Insulator on Transmission Line Based on SSD Algorithm *Instrument Use* **026**(008): 1-4
[5] Wang W, Wang Y, Han J and et al. 2017 Recognition and Drop-Off Detection of Insulator Based on Aerial Image *International Symposium on Computational Intelligence & Design*. IEEE
[6] Lecun Y, Bottou L, Bengio Y and et al. 1998 Gradient-based learning applied to document recognition Proceedings of the IEEE 86(11): 2278–2324
[7] Krizhevsky A, Sutskever I and Hinton G E. 2017 ImageNet Classification with Deep Convolutional Neural Networks Commun. ACM 60(6): 84–90
[8] Simonyan K and Zisserman A 2015 Very Deep Convolutional Networks for Large-Scale Image Recognition International conference on learning representations
[9] He K, Zhang X, Ren S and et al. 2016 Deep Residual Learning for Image Recognition 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
[10] Shelhamer E, Long J and Darrell T 2017 Fully Convolutional Networks for Semantic Segmentation IEEE Trans. Pattern Anal. Mach. Intell. 39(4): 640–651
[11] Ronneberger O, Fischer P and Brox T 2015 U-Net: Convolutional Networks for Biomedical Image Segmentation International Conference on Medical Image Computing and Computer-Assisted Intervention (Springer International Publishing) 234-241
[12] Badrinarayanan V, Kendall A and Cipolla R 2017 SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation IEEE Transactions on Pattern Analysis and Machine Intelligence 39(12): 2481–2495
[13] Zhao H, Shi J, Qi X and et al. 2017 Pyramid Scene Parsing Network 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
[14] Chen L C, Papandreou G, Kojjinos I and et al. 2014 Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs arXiv e-prints
[15] Chen L, Papandreou G, I. Kokkinos and et al. 2018 DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs IEEE Transactions on Pattern Analysis and Machine Intelligence 40(4): 834-848
[16] Chen L C, Papandreou G, Schroff F and et al. 2017 Rethinking Atrous Convolution for Semantic Image Segmentation arXiv e-prints arXiv:1706.05587