Progress In Electromagnetics Research C, Vol. 138, 191–203, 2023

Infrared Image Segmentation Method Based on DeepLabV3+ for Identifying Key Components of Power Transmission Line

Donglei Weng[1], Shuliang Dou[2], Haozhe Wang[3], Dawei Gong[3], Qun Wang[1], and Sailing He[3, *]

Abstract—To improve the work efficiency of on-site inspection personnel in diagnosing faults of power transmission lines, in this paper we propose an infrared image segmentation method based on DeepLabV3+ for identifying key components of transmission line. We collected 556 infrared images of transmission lines in our own power supply system, and expanded the original data by data augmentation method. Based on the comparison of the DeepLabV3+ model with three different backbone networks, MobileNetV2 with the best performance is selected as the main backbone network. Compared with FCN, U-Net, and SegNet, the test results show that DeepLabV3+ using MobileNetV2 (compared with ResNet50 and Xception) can segment the five types of key components in power transmission lines from infrared images more accurately and faster. The MIoU on the test set is 0.8624, which is better than the performance of FCN, U-Net, and SegNet. This lays a foundation for improving the work efficiency of on-site inspection personnel and improving the continuous power supply capacity, stability, and safe operation level of the power grid.

1. INTRODUCTION

The normal and safe operation of power equipment is an important factor in the planning and management of power system, and plays a vital role in the operation of power system. The state of power equipment affects the safe and stable operation of power generation, power transformation, transmission, distribution and electricity consumption. Since the equipment is in a long-term high voltage operation state, and long-term exposure to the external environment, due to aging, corrosion [1–3], a power transmission line may show failure modes such as overheating, arc, partial discharge and other conditions. If these failures can not be found and dealt with in time, it will not only affect the stability of the power supply, but also may cause serious safety accidents. Therefore, regular inspection of a transmission line has become an important task for power grid operation and management, which will greatly save operation and maintenance costs and make correct decisions for future development.

The traditional inspection method of a transmission line mainly relies on manual observation and detection. This method has high work intensity, low efficiency, and is susceptible to the subjective factors and working hours of inspection personnel, resulting in misjudgment of the status of a transmission line which means it is difficult to ensure the accuracy and comprehensiveness of inspection. Infrared thermal image technology has been widely used in the inspection of a transmission line. Through the infrared thermal image scanning of equipment, the abnormal thermal state of equipment can be found in time, so as to achieve the prevention of equipment failure. The accurate segmentation of transmission line components from the infrared image is an important operation for fault diagnosis.

Received 19 August 2023, Accepted 18 October 2023, Scheduled 30 October 2023

* Corresponding author: Sailing He (sailing@zju.edu.cn).

[1] State Grid Zhejiang Electric Power Co., LTD., Ningbo Power Supply Company, Ningbo 315000, Zhejiang Province, China. [2] Ningbo Power Transmission Construction Co., LTD., Yongyao Technology Branch, Ningbo 315000, Zhejiang Province, China. [3] National Engineering Research Center for Optical Instruments, Centre for Optical and Electromagnetic Research, Zhejiang University, Hangzhou 310058, China.

At present, the traditional digital image processing methods and deep learning [4, 5] methods are mainly used to segment the infrared image of transmission line components. The traditional digital image processing methods mainly include the method based on edge detection, the method based on region segmentation, the method based on threshold segmentation, etc. Most of these methods are designed for some specific equipment, and largely depend on expert experience. Since the infrared image has the characteristics of complex background, large thermal radiation interference, and not obvious features, traditional segmentation shows poor versatility and can not be applied to the segmentation of various types of components [6–9].

In recent years, with the rapid development of deep learning technology in the field of computer vision and remarkable achievements, it also brings new possibilities to the normal operation of power line system [10–12] and infrared image segmentation. Deep learning is a computational model that mimics the working mechanism of the human brain. By learning from a large amount of data, it can automatically extract and identify the features in the data, thus achieving excellent performance in many tasks. In the problem of image segmentation, deep learning also shows great potential. Long et al. [13] first proposed the Fully Convolutional neural Networks (FCN) in 2015, in which the encoder performs subsampling to extract the features of the input image, and the decoder performs upsampling to gradually recover the details, thus obtaining the segmentation mask map with the same size as the input image. In order to maintain the comprehensive information contained in the original image, FCN replaces the full-connection layer in the previous classification with a full-convolutional layer at the output end to achieve end-to-end prediction at the pixel level. Since then, other improved methods have also been developed, the main difference being in the decoder and decoder architecture, such as U-Net [14] and SegNet [15], which adopts symmetrical structure. The DeepLab series [16–18] uses multi-scale feature fusion technology to introduce void convolution into the backbone network, both enlarging the receptive field and obtaining high-resolution feature maps. DeepLabV3+, which was proposed in 2018, uses the Encoder-Decoder structure to combine the high level semantic features and low level detail features after Atrous Spatial Pyramid Pooling (ASPP), and refined the segmentation edges. The result is a finer segmentation mask map.

This technology has achieved good results in visible image segmentation [19]. Therefore, this paper uses the advantages of deep learning to propose an infrared image segmentation method for key parts of power transmission and transformation equipment based on DeepLabV3+ to segment key components of transmission line under complex background, which can improve the work efficiency of on-site inspection personnel. It provides technical support for improving the continuous power supply ability, stability and safe operation level of the power grid.

2. DATASET COLLECTION AND PROCESSING

2.1. Collection of Infrared Image Dataset of Transmission Line

In this paper, the power transmission line is taken as the research object, and the FLIR thermal imager is used to collect the images when a power grid company detects the substation equipment under its management. The collected infrared image data include those for five types of transmission line components, namely, insulator, wire, tensioning tube, drainage wire, and wire clamp. The image size is 640×512 color pictures, and the number of original sample images is 556. The image segmentation method based on deep learning needs the support of a large amount of data. If a relatively large CNN network is trained with too little data, overfitting may occur. Therefore, in this paper, image enhancement processing such as sample rotation, Gaussian blur, increased noise and mirror symmetry is applied to expand the data to improve the robustness and accuracy of the network model, and the number of expanded samples is 2780. Fig. 1 shows several infrared images randomly selected in the sample set and their enhancement effects. Fig. 1(a) in the first row is the original image, and Fig. 1(b) in the second row corresponds to the enhancement effect of the first row, and its enhancement processing is mirror symmetry.

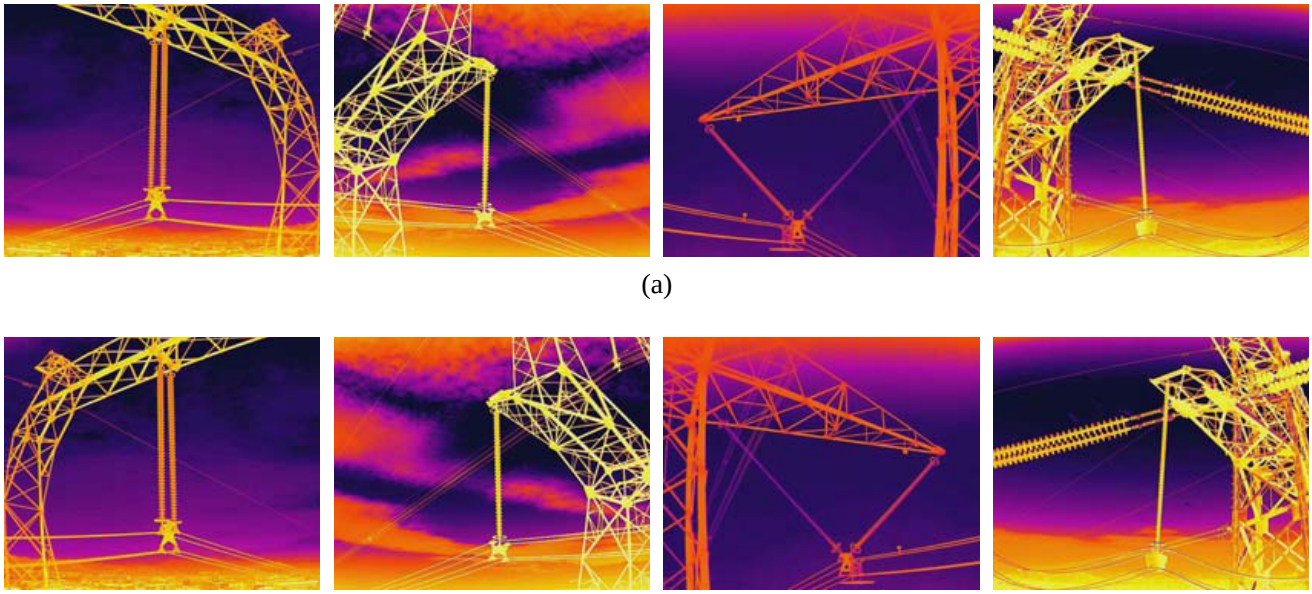


Figure 1. The examples of image augmentation. (a) The original image. (b) The enhanced image.

2.2. Labeling of Datasets

The data collected in this sample are for five types of power transmission line components: insulators, conductors, tensioning tubes, drainage wires and wire clips. Therefore, insulators, conductors, tensioning tubes, drainage wires and wire clips were taken as the segmentation targets, and labeled with Labelme respectively. Specifically, in this paper, the insulator, wire, strained tube, drain wire and clamp in the image are labeled with polygons with category names when labeling. The tagging

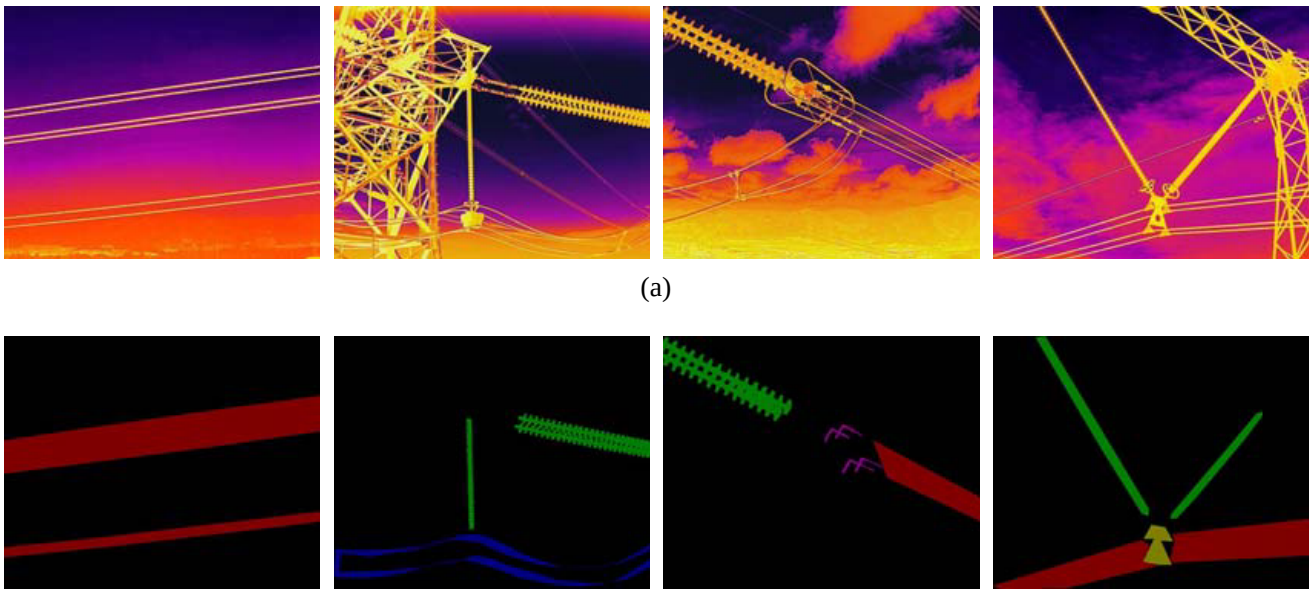


Figure 2. Original images in data with labels. (a) The original image. (b) Ground truth.

example is shown in Fig. 2. The tagging of the target forms the Ground Truth of the divided image, and generates a json file. In the tagging image, red represents the wire, green represents the insulator, blue represents the drainage wire, purple represents the strained tube, and yellow represents the clamp.

2.3. Division of Dataset

In order to better evaluate the network performance and verify the generalization of the model, the above data sets are divided. As shown in Table 1, 70% of the data is divided into training sample sets for training neural network models. 15% is divided into verification set, which is used to verify the training effect and make hyperparameter adjustment, and 15% as a test set, used to test the generalization ability of the model.

Table 1. Partition distribution of the dataset.

Total number of data sets	Image size	Number of training sets, verification sets, and test sets
2780	640 × 512	Training set: 1946 Verification set: 417 Test set: 417

3. NETWORK STRUCTURE

3.1. DeepLabV3+ Network

The DeepLabV3+ network structure used in this paper is shown in Fig. 3, which adopts the Encoder-Decoder structure. The main function of the encoder is to extract useful features from the input

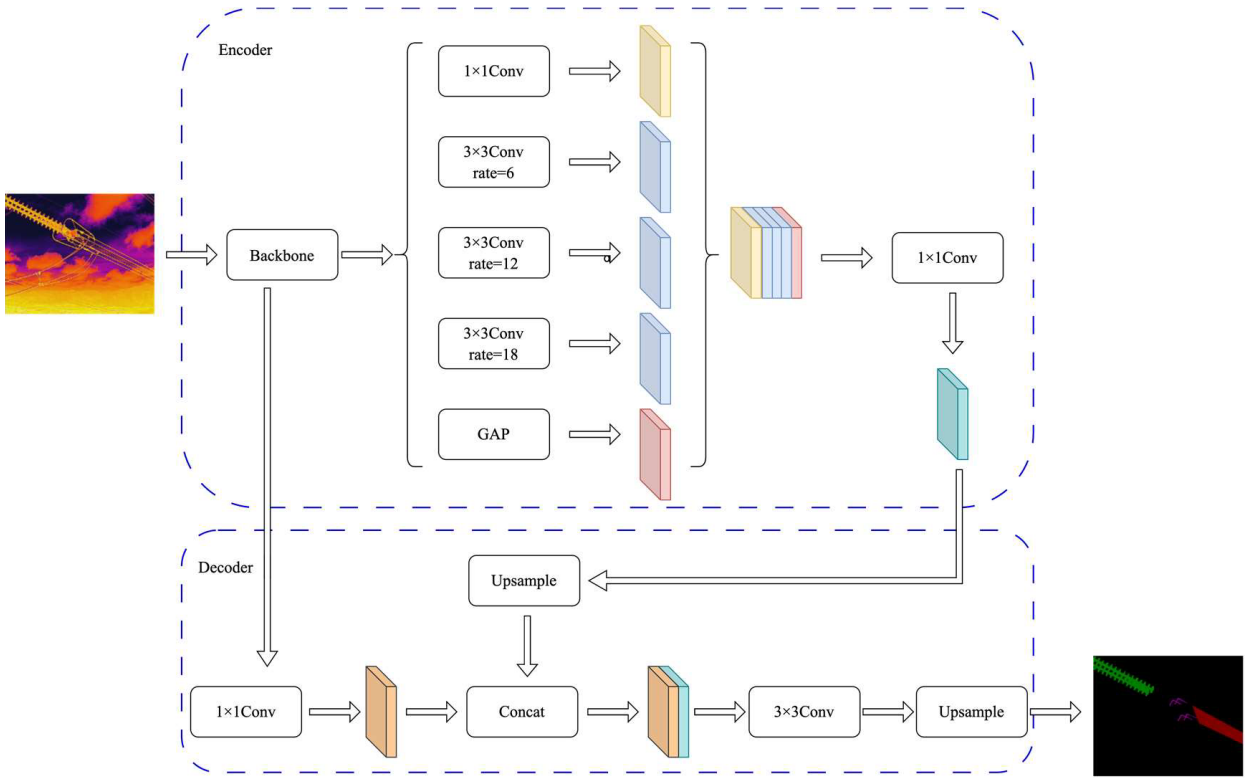


Figure 3. DeepLabV3+ network.

image by using the neural network with image-level classification capability. The decoder part is mainly responsible for upsampling by bilinear interpolation, recovering the spatial information from the features extracted by the encoder, and finally generating the segmentation mask map with the same size as the input image. The input image first passes through the encoder part of the backbone network Xception, and then through the empty space convolution Pyramid Pooling structure (ASPP), ASPP structure uses void convolution of different expansion rates and Global Average Pooling (GAP) to obtain multi-scale information on the extracted feature map of the backbone network, and then uses 1×1 convolution for feature fusion and is processed by the decoder. In the decoder, the decoder first fuses the shallow features extracted from the backbone network with the deep features after 4x up-sampling by ASPP module, then extracts the feature information after the fusion through 3×3 convolution, and finally performs 4x up-sampling on the processed features to output the segmentation mask map with the same size as the input image.

3.2. Xception Backbone Network

Xception [20] is another improvement on Inception V3 proposed by Google after Inception. Xception mainly uses Depthwise Separable Convolution to replace the convolutional operations in Inception V3. A residual structure is introduced to solve the gradient disappearance and explosion problems caused by

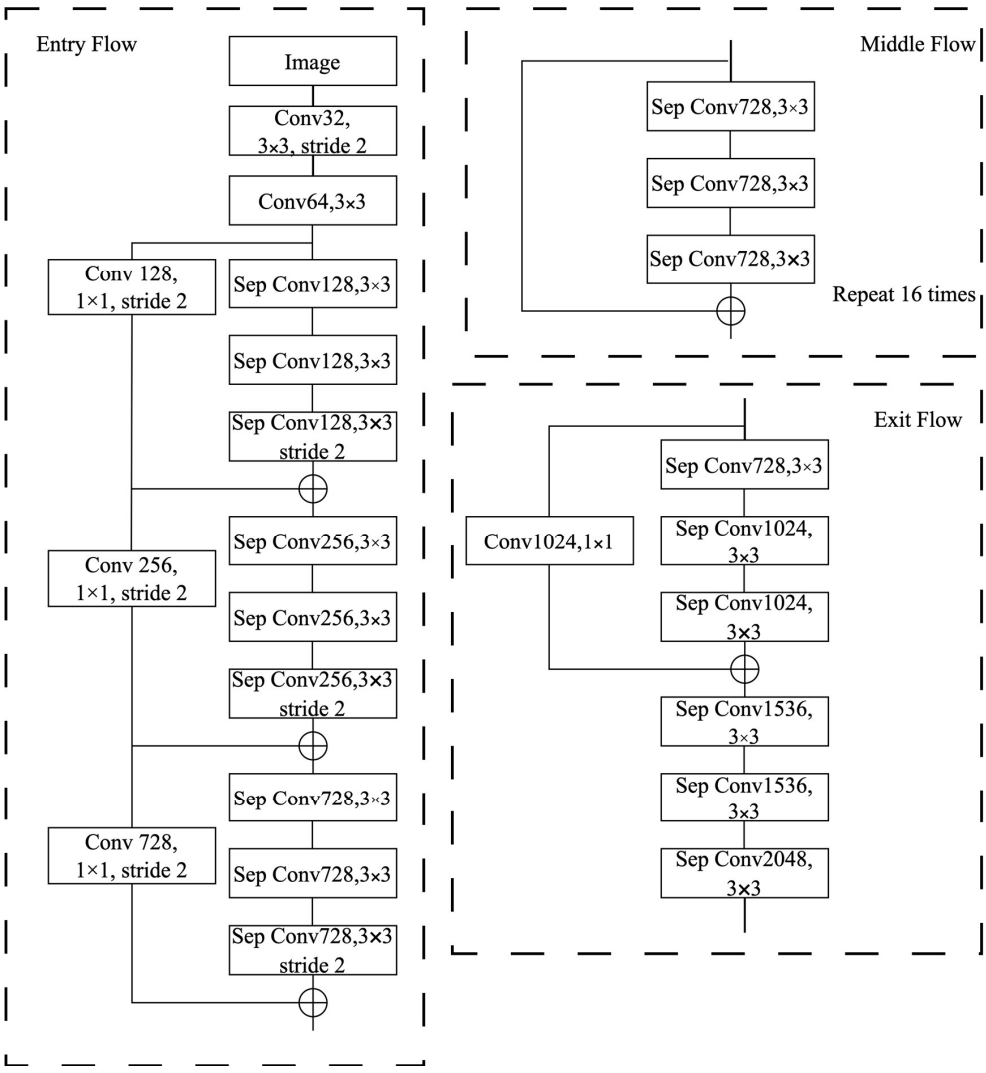


Figure 4. Modified Xception structure.

excessive depth of the network. Depth-separable convolution is used for model compression to reduce the number of model parameters. The final model achieves better results on ImageNet and other datasets than Inception V3, and its model size and computational efficiency are also greatly improved compared with Inception V3. Inspired by MSRA team-aligned Xception, the Xception used in Deeplabv3+ has some changes compared to previous Xception: The first is a deeper structured Xeception; The second is that all Max-pooling layers are replaced by a 3×3 deep separable convolution with step 2, and the last is the addition of Batch Normalization and activation function Relu after the 3×3 deep separable convolution. The modified Xception structure is shown in Fig. 4.

3.3. MobileNetV2 Backbone Network

MobileNetV2 [21] is a lightweight deep neural network architecture designed for efficient computing on devices with limited resources, such as mobile and embedded devices. It was proposed by Google in 2018 as a successor to MobileNetV1, and the network exhibits more accurate model performance while maintaining low computational complexity.

The main innovations in MobileNetV2 are the introduction of Inverted Residuals and Linear Bottlenecks. In MobileNetV2, information is transferred in the opposite way to traditional residuals networks. Information is first convolved through a wide convolutional layer, then through a narrow Bottleneck, and finally through a wide convolutional layer. The main idea of this design is that the nonlinear transformation should be carried out in the low-dimensional space, while in the high-dimensional space, the network should keep the original information as much as possible. In addition, the activation function of the last convolution layer of each residual block is removed, making this layer a linear transformation. The purpose of this design is to prevent nonlinearities from destroying too much information by nonlinear activation functions such as ReLU that might destroy the feature's primitiveness.

This design enables MobileNetV2 to achieve high model performance while maintaining computational efficiency. As a result, MobileNetV2 has become a popular choice in many application scenarios where computing resources are limited, including tasks such as image classification, object detection, and semantic segmentation.

3.4. Atrous Spatial Pyramid Pooling (ASPP)

Atrous Spatial Pyramid Pooling (ASPP) [11] module contains a series of empty convolution with different expansion rates. Its main function is to use these empty convolution with different expansion rates to extract multi-scale context information and integrate the extracted multi-scale information together. To achieve the purpose of covering multi-scale receptive field, which is very useful for dealing with objects of different sizes. The structure of ASPP is shown in Fig. 5. ASPP consists of one convolution kernel of size 1×1, three expansion convolution cores of size 3×3 with expansion rates of 6, 12 and 18 respectively, and one Global Average Pooling (GAP). In common convolutional networks, to increase the receptive field, often requires an increase in the depth of the network or a large convolutional kernel, which will increase the computational complexity and the number of parameters of the network. In contrast, ASPP, by using empty convolution, can effectively enlarge the receptive field so that each convolutional output contains a larger range of information without increasing the computational complexity and number of parameters. The larger the expansion rate, the larger the feature range extracted, the larger the receptive field generated, and therefore the more semantic feature information obtained in the image.

4. EXPERIMENTAL PROCESS, RESULTS AND ANALYSIS

4.1. Experimental Environment

The operating system of the experimental environment in this paper is Ubuntu 16.04, the processor is Intel(R) Xeon(R) Silver 4110 CPU, the memory is 128 GB, and the graphics card is NVIDIA GeForce RTX 2080Ti, version 525.60.11. CUDA version 10.1 and cuDNN version 8.0 are adopted. The network

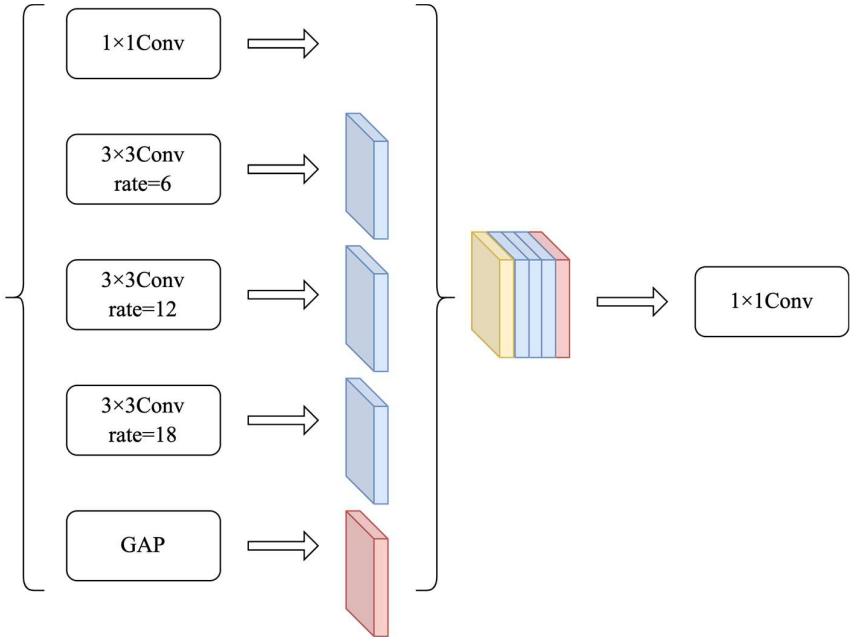


Figure 5. ASPP structure.

models involved are all implemented using the Pytorch deep learning framework, and the programs are written using Python3.7 in PyCharm.

4.2. Implementation Details

In the training model stage, the processed data set is tested. The optimizer selects the Adam optimizer, the initial learning rate is 0.0005, the momentum term is 0.9, and the weight attenuation coefficient is 0.0001. The learning rate is adjusted by step_size = 10 and the multiple gamma = 0.92. The batch training size Batchsize = 8, the number of iterations (epoch) is 300, and the loss function is the cross entropy loss function.

4.3. Evaluation Metrics

In order to evaluate the performance of the model, Intersection over Union (IoU) and Mean Intersection over Union (MIoU) were used as evaluation indexes to compare and analyze the test results. The Intersection over Union (IoU) represents the ratio between the intersection and union of two sets of predicted and true values of the model for a certain class. The calculation formula is shown in (1). The average intersection ratio represents the average of the intersection and union ratio between the predicted value and the real value set, and can reflect the degree of overlap between the segmentation result and the real label. The calculation formula is shown in (2).

$$\mathrm{IoU}_i = \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \tag{1}$$

$$\mathrm{MIoU}_i = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \tag{2}$$

where k is the total number of pixel classification categories; p_{ij} and p_{ji} respectively represent the total number of pixels predicted as i and the total number of pixels predicted as j and the total number of

pixels predicted as j and the actual result as i, while p_{ii} represents the predicted result as i, and the real result is also the total number of pixels i.

4.4. Analysis of Results

4.4.1. Comparison of Different Backbone Networks

To achieve the optimal segmentation performance of DeepLabV3+, ResNet50, MobileNetV2, and Xception are respectively used as backbone feature extraction networks for training in the experiment, so as to compare performance. Their loss change curves are shown in Fig. 6. As can be seen from Fig. 6, Xception-DeepLabV3+ model converges when Epoch is about 300, MobileNetV2-DeepLabV3+ model and ResNet50-DeepLabV3+ model basically converges when epoch is about 100, and the convergence speed is fast. And the MobileNetV2-DeepLabV3+ model has the lowest loss value. Therefore, using MobileNetV2 as the backbone feature network has fast convergence speed and high precision.

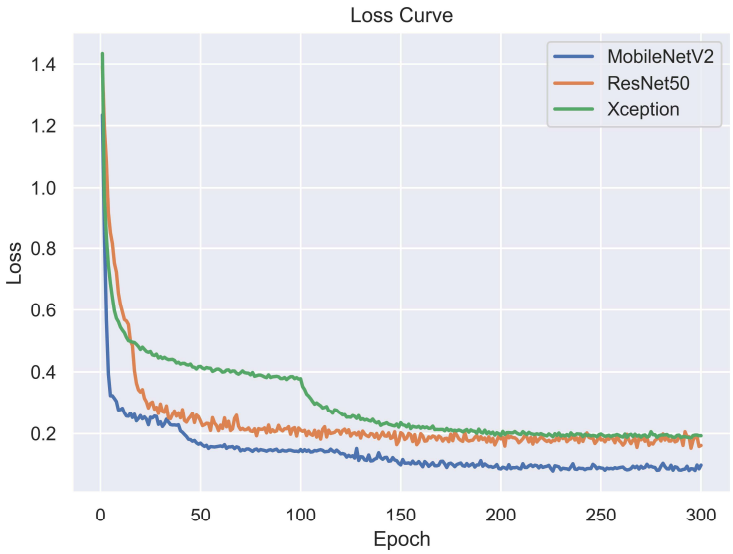


Figure 6. Comparison of training losses of different models.

The test results of three different backbone feature extraction networks, Resnet50, MobileNetV2, and Xception, on DeeplabV3+ are shown in Table 2. When MobileNetV2 is used as the backbone feature network, the DeepLabV3+ model has the best performance in dividing insulator, wire, strained tube, drain wire and clamp, and its crossover ratio can reach 0.86, 0.9043, 0.7295, 0.8011, 0.8941, respectively, and IoU can reach 0.8624. Are better than ResNet50 and Xception. The experimental results show that the DeepLabV3+ model method of MobileNetV2 has the best comprehensive performance for the key components segmentation task of transmission line in this paper.

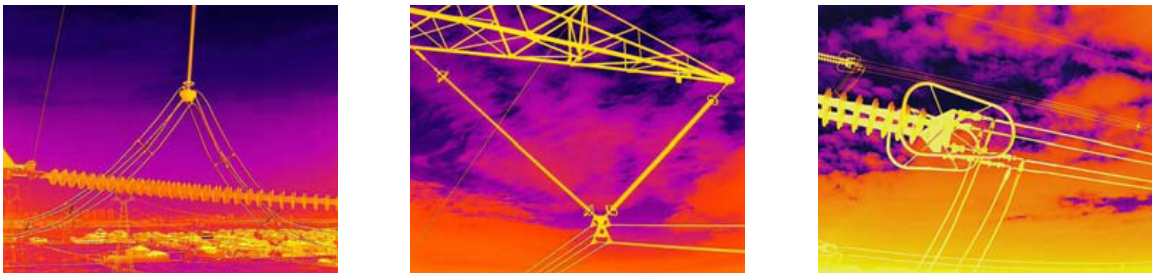
4.4.2. Comparison of Different Segmentation Networks

In order to facilitate the comparison experiment with other commonly used deep learning semantic segmentation networks, FCN, U-Net, and SegNet semantic segmentation networks are implemented under the same environment, and the same training set and test set are used to train and test the model.

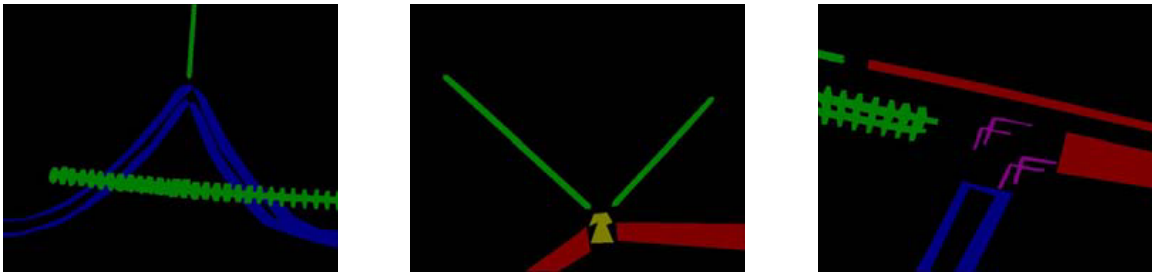
Figure 7 shows the output prediction results of different network models, where (a) represents the test original; (b) represents the label corresponding to the test original; (c) represents the prediction result of FCN model; (d) represents the SegNet model; (e) represents the U-Net model; and (f) represents the prediction result of DeepLabV3+ model. All the four models can predict the spatial position and partial outline of the device from the complex background, but it is still inevitable that some pixels

that do not belong to the five categories of components such as insulators and drainage wires will be classified into this category, while those belonging to the five categories such as insulators and drainage wires will be predicted as the background. Among them, the FCN model has a higher degree of false segmentation, followed by the SegNet model. The prediction result of U-Net can be close to the label of the test image to a greater extent, but compared with DeepLabV3+ model, DeepLabV3+ model has better control over the details of image segmentation, and its shape edge is clearer.

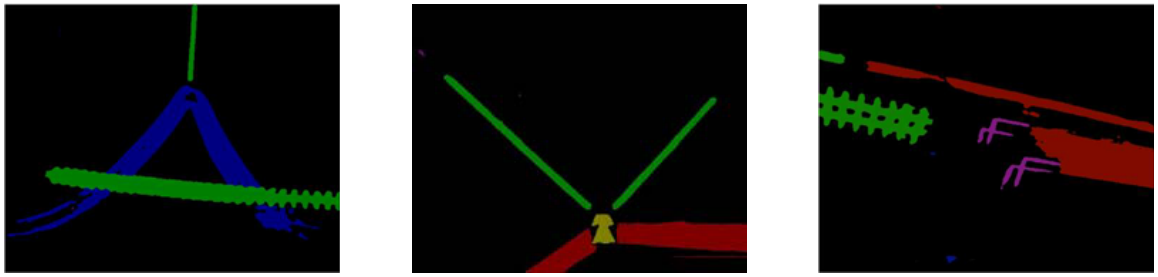
Specifically, FCN performs well for component segmentation in a simple background, and can accurately segment components such as insulators and clamps. However, it performs poorly in complex backgrounds (susceptible to noise), and is not accurate enough for the segmentation of smaller targets such as wires and drainage wires. The encoder-decoder structure of U-Net helps to capture detailed information, and has a good segmentation effect for insulators and wires. However, its segmentation performance for drainage wires in complex background is poor, and more data are needed for training. SegNet uses an encoder-decoder structure similar to U-Net, which can effectively segment insulators, clamps, etc., while having low computational complexity. However, the segmentation performance



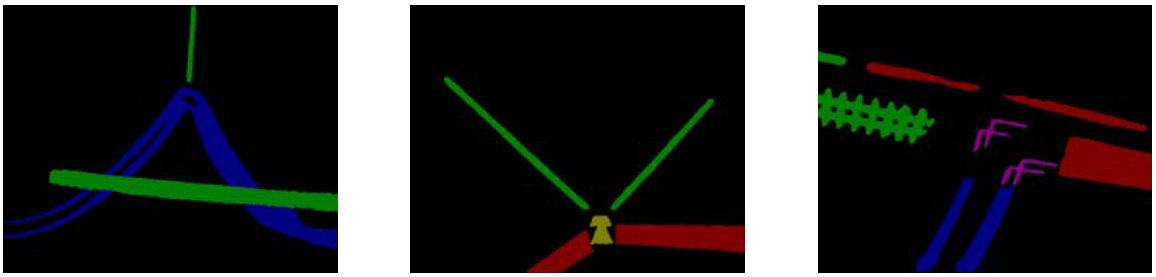
(a)



(b)



(c)



(d)

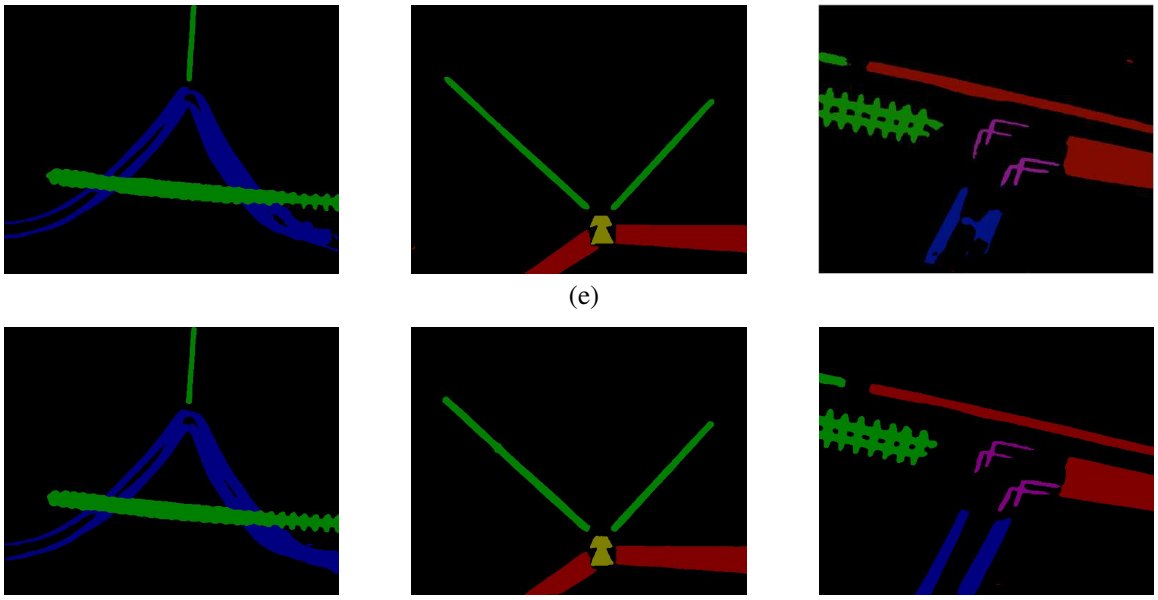


Figure 7. The output of DeepLabV3+ and different models. (a) Original image. (b) Ground truth. (c) FCN. (d) SegNet. (e) U-Net. (f) DeepLabV3+.

Table 2. Comparison of experimental results of different models.

Models	Categories	IoU	MIoU
ResNet50	Insulator	0.8324	0.8246
	Wire	0.8616	
	Strained tube	0.676	
	Drain wire	0.7285	
	Clamp	0.8695	
	Background	0.9794	
Xception	Insulator	0.8197	0.8075
	Wire	0.8299	
	Strained tube	0.6580	
	Drain wire	0.6932	
	Clamp	0.8685	
	Background	0.9758	
MobileNetV2	Insulator	0.86	0.8624
	Wire	0.9043	
	Strained tube	0.7295	
	Drain wire	0.8011	
	Clamp	0.8941	
	Background	0.9853	

is poor in complex background. DeepLabV3+ performs well in simple computing environments and performs better than other models for the segmentation of these five parts. It may have limited performance in segmenting some details in complex backgrounds, but still performs the best as compared to other models.

Table 3 shows the FCN, U-Net, and SegNet methods as well as the evaluation indexes of DeepLabv3+ method in the test data. The test results of the test set show that the IoU and MIoU of the five types of transmission line component (insulator, wire, strained tube, drain wire and clamp) and background using the DeepLabv3+ method proposed in this paper are higher than those of the FCN, U-Net, and SegNet methods.

Table 3. Comparison of experimental results of different segmentation models.

Models	Categories	IoU	MIoU
FCN	Insulator	0.8141	0.7755
	Wire	0.7462	
	Strained tube	0.6555	
	Drain wire	0.5929	
	Clamp	0.8787	
	Background	0.966	
U-Net	Insulator	0.8437	0.8435
	Wire	0.8534	
	Strained tube	0.7230	
	Drain wire	0.7768	
	Clamp	0.8831	
	Background	0.9812	
SegNet	Insulator	0.843	0.8347
	Wire	0.8463	
	Strained tube	0.7056	
	Drain wire	0.7569	
	Clamp	0.8761	
	Background	0.9801	
DeepLabV3+	Insulator	0.86	0.8624
	Wire	0.9043	
	Strained tube	0.7295	
	Drain wire	0.8011	
	Clamp	0.8941	
	Background	0.9853	

5. CONCLUSION

In this paper we take power transmission lines as the research object, and study the object segmentation of infrared image of transmission lines under complex background, and construct DeepLabV3+ network model to segment the infrared image of key components intransmission lines. The model is tested on the data set of transmission line and compared with FCN, U-Net, and SegNet. The experimental results show that: The DeepLabv3+ method proposed in this paper can segment the five types of transmission line component, namely insulator, wire, strained tube, drain wire and clamp, from the infrared image more accurately. It can effectively sort the insulator, wire, strained tube, drain wire, clamp and background in the image of transmission line pixel by pixel, and the final MIoU on the test set is 0.8624. The IoU and MIoU of the five types of transmission line components and the background are higher than those of FCN, U-Net, and SegNet methods. This provides technical support for improving the work efficiency of on-site inspection personnel, and improving the continuous power supply ability, stability and safe operation level of the power grid. A limitation of the present study is that we used

a dataset of limited size and limited diversity. Future work includes expanding the dataset size and ensuring the inclusion of more categories and scenarios to improve the generalization performance of the model. Another limitation of the work is that we chose a particular model architecture in our study, but there are other architectures that may be more suitable and may have room for the improvement of the performance. We plan to try different architectures in the future to determine which is the best for our task.

ACKNOWLEDGMENT

The work is partially supported by "Pioneer" and "Leading Goose" R&D Program of Zhejiang (2023C03135), Ningbo Science and Technology Project (2021Z029), and National Natural Science Foundation of China (11621101). The authors are grateful to Dr. Julian Evans of Zhejiang University for valuable discussions.

REFERENCES

1. Zhao, Z., N. Liu, and L. Wang, "Localization of multiple insulators by orientation angle detection and binary shape prior knowledge," *IEEE Transactions on Dielectrics and Electrical Insulation*, Vol. 22, No. 6, 3421–3428, 2015.
2. Tao, X., D. Zhang, Z. Wang, et al., "Detection of power line insulator defects using aerial images analyzed with convolutional neural networks," *IEEE Transactions on Systems Man and Cybernetics Systems*, Vol. 50, No. 4, 1486–1498, 2020.
3. Lin, T. and X. Liu, "An intelligent recognition system for insulator string defects based on dimension correction and optimized faster R-CNN," *Electrical Engineering*, Vol. 103, 541–549, 2021.
4. Gong, D., T. Ma, J. Evans, and S. He, "Deep neural networks for image super-resolution in optical microscopy by using modified hybrid task cascade U-Net," *Progress In Electromagnetics Research*, Vol. 171, 185–199, 2021.
5. Shou, Y., Y. Feng, Y. Zhang, H. Chen, and H. Qian, "Deep learning approach based optical edge detection using ENZ layers," *Progress In Electromagnetics Research*, Vol. 175, 81–89, 2022.
6. Zhang, S. C., Y. L. Xu, N. Yang, et al., "Insulator image extraction algorithm based on maximum entropy and Hough transform," *East China Electric Power*, Vol. 40, No. 4, 595–598, 2012.
7. Zhang, Z. J., Y. Zhang, X. L. Jiang, et al., "A model for predicting ice weight on insulators based on regional segmentation method," *High Voltage Technology*, Vol. 46, No. 2, 406–412, 2020.
8. Liu, Z. T., J. G. Yin, K. D. Li, et al., "Infrared image segmentation method of insulator string based on watershed algorithm," *Electric Porcelain Arrester*, Vol. 228, No. 2, 216–221, 2020.
9. Li, H. R., J. Gao, T. Wu, et al., "Research on insulator crack detection based on improved Canny operator," *Smart Electric Power*, Vol. 49, No. 2, 91–98, 2021.
10. Li, W. P., K. Xie, X. Liao, et al., "Intelligent diagnosis method of infrared image for transformer equipment based on improved faster RCNN," *Southern Power System Technology*, Vol. 13, No. 12, 79–84, 2019.
11. Zhou, K. H., Z. W. Liao, Y. Y. Xiao, et al., "Construction of infrared image classification model for power equipments based on improved CNN," *Infrared Technology*, Vol. 41, No. 11, 1033–1038, 2019.
12. Jia, X., J. L. Zhang, and X. B. Wen, "Infrared faults recognition for electrical equipments based on dual supervision signals deep learning," *Infrared and Laser Engineering*, Vol. 47, No. 7, 0703003-1–7, 2018.
13. Long, J., E. Shelhamer, T. Darrell, et al., "Fully convolutional networks for semantic segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440, 2015.
14. Ronneberger, O., P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, 234–241, 2015.

15. Badrinarayanan, V., A. Kendall, and R. Cioplla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 12, 2481–2495, 2017.
16. Chen, L. C., G. Papandreou, I. Kokkinos, et al., "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 40, No. 4, 834–848, 2018.
17. Xiong, F.-G., X. Zhang, X. Han, et al., "Improved semantic segmentation of remote sensing image," *Computer Engineering and Applications*, Vol. 58, No. 8, 185–190, 2022.
18. Chen, L. C., Y. Zhu, G. Papandreou, et al., "Encoder-decoder with atrous separable convolution for semantic image segmentation," *Proceedings of the European Conference on Computer Vision*, 801–818, 2018.
19. Wang, C., "Research on infrared image semantic segmentation based on deep learning," Ph. D. Dissertation of University of Chinese Academy of Sciences (Shanghai Institute of Technical Physics, Chinese Academy of Sciences), Shanghai, 2017.
20. Chollet, F., "Xception: Deep learning with depthwise separable convolutions," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1800–1807, 2017.
21. Sandler, M., A. Howarda, M. L. Zhu, et al., "MobileNetV2: Inverted residuals and linear bottlenecks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4510–4520, 2018.