UAV Transmission Line Inspection Object Recognition based on Mask R-CNN

Yi Liu, Haolin Huo, Jiefeng Fang, Junting Mai, Shengxiang Zhang*
College of Mathematics and informatics South China Agricultural University
Guangzhou, China, 510642
*Corresponding author e-mail: sxzhang@scau.edu.cn

Abstract. Currently, unmanned aerial vehicles (UAV) are applied to routine inspection tasks of power transmission devices. Deep-learning algorithm and machine vision have attracted much attention in the field of the UAV’s autonomous control as it’s an effective way to improve the efficiency of inspection. Considering the differences between the distant and close view, this paper adopts Mask R-CNN to detect various components of power transmission devices but use the methods such as processing of the edge, hole filling and Hough Transform identify the wires in distant. Some major components, such as pole, truss, cross arm, insulator string and so on, can be 100% recognized. This proposed model shows the characteristics of high recognition speed and high accuracy, which can assist UAV to inspect well.

1. Introduction
In recent years, with the gradual expansion of China's power grid laying area and transmission line construction scale, it has gradually covered areas with harsh environments and complex areas. In order to maintain the reliability, availability, and sustainability of the power supply, power companies regularly inspect the transmission equipment. Traditional methods of checking the grid usually include field surveys and air surveys. The core methods are mainly manual inspections, such as walking groups or helicopter teams, which have not changed for decades [1]. As manual inspection cannot meet the growing business, UAV power inspection has become a major inspection mode, the drone gradually automated and efficient electrical safety inspection by remote sensing technology and sensors.

In order to avoid the collision of UAV and implement navigation control, and to better improve the efficiency of inspection, it is very necessary to analyze and screen the inspection contents in combination with computer vision. In computer vision, with the rapid development of deep learning, the accuracy of object detection has made great progress.

Next, we will present the current mainstream image recognition and instance segmentation techniques from the existing theoretical aspects of deep learning, and then apply them to our datasets to show the test results.

2. Study area and imagedata

2.1. Study area
So far, there has been a lot of research on the inspection of transmission equipment with deep learning combined with UAV. Xiaolong Hui and Jiang Bian combine the deep learning method of Faster R-CNN and YOLOv2 with fast-tracking to realize UAV navigation based on transmission platform, which has
achieved good results in image detection [2]. However, there is no literature on the identification of all parts of the transmission lines.

Transmission line detection is one of the most important steps of UAV power inspection. Hough transform [3] is a widely used method. Tang Wen YANG et al. first converted the color camera image into a grayscale image and converted it into a binary image by a threshold method. The Hough transform is introduced to detect the candidate rows in the binary edge image. Finally, the fuzzy c-means clustering algorithm is used to discriminate the power line and the detected candidate lines. The algorithm is effective for automatic detection of power lines and can withstand complex terrain backgrounds and various illumination noises [4], but for other non-topographic complex backgrounds, there is no uniform, general-purpose algorithm research.

2.2. Image data

The image data of this article comes from the Shanxin lines taken by the Maoming Bureau of Guangdong Province, which was taken by UAV. The inspection method is visible light, and the voltage level is 110kV. There are so many backgrounds in these images, such as mountains, gardens, houses, trees, lakes and so on. The size of these images is 8688×5792, and in this paper, we compressed them to 2560×1792 for labeling and training.

3. Methodology

3.1. Deep learning

Deep learning is a new field in machine learning research. Multilayer perceptron with multiple hidden layers is a deep learning structure. The motivation lies in the establishment of a neural network that simulates the human brain for analytical learning and combines low-level features to form more abstract high-level representation attribute categories or features to discover distributed features of data.

Deep learning allows computational models consisting of multiple processing layers to learn data representations with multiple levels of abstraction that greatly improve the skill level of speech recognition, visual target recognition, target detection, and many areas such as drug discovery and genomics. Deep convolutional networks have made breakthroughs in the image, video, speech, and audio processing, while periodic convolutional networks have brought light to sequence data such as text and speech [5].

In recent years, deep learning has attracted more and more people's attention. Long et al. introduced a fully convolutional network, applied neural networks to image classification, and formed a fully convolutional network to semantically segment images [6]. A recent variant of FCNs is based on a masked region of the Convolutional Neural Network (Mask R-CNN) [7], which enables instance segmentation. The instance split generates a bounding box, a class label, and a pixel-precision split mask for each OOI in the image. Mask R-CNN is able to achieve impressive results.

3.2. Mask R-CNN

Conceptually, Mask R-CNN [8] is easy to understand. Branches based on candidate object class labels and bounding boxes, which is the Faster R-CNN [9] output while Mask R-CNN adds the third branch that output object mask. The output of the mask needs to extract the geometric position of the finer object, which is different from the output of the category and border. To this end, Mask R-CNN uses a technique that plays a key role in the network, which is to replace RoIPool's operation with RollAlign. Mask R-CNN uses a two-stage strategy that inherits the idea of Faster R-CNN. In the first stage, a recommendation area is generated for the RPN(Region Proposal Network) network, and the RoI(Region of Interests) is mapped to the feature map by bilinear interpolation. The main content of the second stage is to predict the category and frame position of the candidate area, different from the Faster R-CNN is that the candidate region which Mask R-CNN join in corresponds to the branch of the object's binary mask. In particular, the category and border positions are calculated in parallel during the inspection
process. In addition, the mask of the RoI region is not dependent on class detection. Therefore, the loss function of the RoI region should be the formula (1).

$$L = L_{cls} + L_{box} + L_{mask}$$  \hspace{1cm} (1)

For an RoI associated with ground-truth class, k is only defined on k-th mask [8]. Such design avoids competing with the detection of the category when calculating the mask.

The generation of the mask requires the extraction of the boundary information of the object, which requires relatively accurate geometric information. However, RoIPool is a quantization operation, the geometric position information of objects in the image is lost during the pooling operation. This is caused by the nature of the quantization operation itself. In order to preserve the geometric position information of objects in the original image as much as possible, Mask R-CNN adopts the method of bilinear interpolation [10]. Then, the feature map of the aligned RoI is input into the FCN (Fully Convolutional Networks) network to implement image segmentation [11].

According to our experience using a deeper network, the network is more possible to learn more details and get much more features from the original images. In this paper, we used ResNet101-FPN backbond to generate a feature map. In Mask R-CNN [8], the result tells us that using a ResNet-FPN backbond for feature extraction gives an excellent gain in both accuracy and speed. This thought inherited from Faster R-CNN.
1. Using ResNet-C4 backbone, which closely follows the architectures of Faster R-CNN.
2. Using FPN backbone (include ResNet), which can detect a smaller object. The feature maps from the ResNet in different stages will be put into the FPN. And the FPN can allocate a set of anchors to different feature maps. For a smaller feature map, the FPN can scale a small region on it, so that the network can detect a smaller object.

![Figure 4](image)

Figure 4. Feature maps from ResNet C4 and feeds to FPN

3.3. Wire
For the reason that wires are slender, the color also changes with the light. Based on complex and numerous transmission lines, it is complicated to process the various background. The study of the picture found that in the close-up picture, the wires are close to the insulator strings and the vibration dampers. The style is relatively thick so that we could use polygon annotations to identify them by using the Mask R-CNN network.

In the distant view, we inspired by the literature [12], and found the breakthrough is the color of the wire. The color of the wire generally close to white, we extracted them and set black while the rest set white. The popular detection extraction method is that the image enhancement processing is first performed, then the power line edge is quickly detected, and finally, the power line is quickly fitted and connected to obtain the final result [13]. On this basis, we found that using enhancement processing did not work well. So we used the canny edge detection and the hole filling which could eliminate most of the background. Then we use the Hough Transform to detect straight lines. The introduction of the principle of wire processing will be divided into three parts.

3.3.1. Canny edge detection. There are four steps for canny edge detection.
1. Gaussian filtering of grayscale images
2. Find the amplitude and angle images of the gradient
3. Non-maximum suppression of the gradient magnitude pattern and double threshold processing.
4. Connect analysis to detect and join edges.

3.3.2. Holes filling. The area in the binary image enclosed by 1 is called a hole. If there is such a region, then set it to 1.

By removing the filled holes from small areas and large areas, the disturbing background can be eliminated while the straight lines can be highlighted.

\[
\begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 \\
1 & 0 & 0 & 1 \\
1 & 1 & 1 & 1 \\
\end{pmatrix} \Rightarrow
\begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 \\
\end{pmatrix}
\] (2)

3.3.3. Hough transform. The Hough Transform is a feature detection that can be used to identify features such as lines, ellipses, and circles. Based on a good understanding of the shape information, it
can be used to find out the shape's position and orientation in the image. In this paper, the main application is the detection of straight lines by using the Hough Transform.

The essence of the Hough Transform is to find all the points in a linear function.

Suppose a straight line equation in a Cartesian coordinate system is shown in equation 3. In the equation, $k$ is the slope and $b$ is the intercept. A point on the line is converted to the polar coordinate system by Hough transform.

$$y = kx + b$$  \hspace{1cm} (3)

At each pixel on this image, find the $(\theta, \rho)$ that satisfies equation 4 in the parameter space. Then count the logarithm of the same value.

$$x \cos \theta + y \sin \theta = \rho$$  \hspace{1cm} (4)

At last, Finally, take out the parameters larger than the set value.

3.4. Implementation

In the beginning, we compress the drone image and determine what objects should be labeled. The 13 objects' pictures which we have determined 13 as figure 2.

In this paper, the open source software package based on Keras and Tensorflow is used, and the Mask R-CNN method is implemented by using the drone to predict the picture which with the polygon as the training set. During the training, a small processor (GeForce GTX 1080ti graphic card) was used to train each small batch with Resnet-101. The parameter settings are as table 1.

| Learning Rate | 0.0001 |
|---|---|
| Figure Size | 640×448 |
| Training Set | 284 |
| Testing Set | 33 |
| Epoch | 30 |
| Train Steps | 950 |
| Validation Steps | 50 |

![Pole](image1.png) ![Cross arm](image2.png) ![Insulator string](image3.png)

![Suspension clamp](image4.png) ![Strain clamp](image5.png) ![Yoke plate](image6.png)

![Cat-like line tower](image7.png) ![Tangent tower](image8.png) ![Transmission tower](image9.png)

![Truss](image10.png) ![Vibration damper](image11.png) ![Grading ring](image12.png) ![Hold hoop](image13.png)

Figure 5. Visual image
Next is the extraction process for the distant wires. Set the RBG value of the pixel color to be extracted (225, 225, 225), and the color to allow the drift range to be 0.5. Taking the input image as 2560×1792 as an example, a small area of fewer than 400 pixels is removed and a large area of more than 5000 pixels is removed. The setting value in the Hough transform is 5. Insert paragraph text here.

4. Result and discussion

4.1. Result
Training based on the above settings. In the course of training, the trend of loss is shown in Figure.

After the loss has bellowed 0.2, we forecasted the pictures. In the reason that the number of the object is large, a forecasted picture could not reflect all the recognition effect of items. So we randomly extracted a picture of the non-test set and the training set, the predicted result is shown in figure 5. As can be seen, most of the objects can be accurately identified. But can also be seen that the strain clamp has been missed. We guessed the reason may be that there were fewer training pictures about it.

We can know that the effect of line recognition is generally through the figure 8. But compared with the traditional method, this is a great improvement.
4.2. Discussion

In the traditional model, the method of transmission line identification such as the recognition of foreign object based on Hough Transform [14], Canny edge detection and SVM identification [15]. These traditional models have many drawbacks and are not suitable for image detection in UAV. The traditional model has high-quality requirements for images. If the background of the image contains many objects characterized by straight lines, this will affect the detection of the wires in the case where the Hough Transform is sensitive to the lines in the image. In particular, the data set used in this paper is in the mountain area, where is difficult to take photos manually, so they are taken by UAV. These images have a variety of backgrounds, and some backgrounds have components such as plants, wood, trees, etc. that contain a large number of straight edges. Therefore, the model should have good generalization ability and accurate recognition ability.

In the SVM-based high-voltage transmission line inspection robot obstacle recognition research and application [15], the author mentioned the use of multiple SVM to detect different fixtures that need to find the best fittings through the coding training form, so it causes a problem of large calculations and long training process. The Mask-RCNN model improves the speed of training greatly and is superior to the traditional SVM classification in multi-sample identification.

The method used in this paper, Mask R-CNN which uses the backbone network to train ResNet-101 to get APs under different IOU Thresholds (Table 2).
Table 2. The average precision

| backbone         | AP75 | AP90 |
|------------------|------|------|
| Mask R-CNN       | 0.375| 0.354|
| ResNet-101-FPN   |      |      |

5. Methodology

Labeled the parts of the transmission line and identified them with Mask R-CNN. Combined with the identification and processing of the wires, the close-up uses network identification. In the distant view, the wires are stripped off by the difference in color between the wires and the surrounding background pixels, effectively eliminating the influence of the foreground-background. Through the network prediction, the parts of the transmission line and the effect of wire identification are excellent. This result can help UAV to inspect transmission lines.

References

[1] Matikainen L, Lehtomäki M, Ahokas E, Hyypää J, Karjalainen M, Jaakkola A, Kukko A, Heinonen T, et al. Remote sensing methods for power line corridor surveys. ISPRS J Photogramm Rem Sens, 2016; 119:10–31.
[2] Xiaolong Hui and Jiang Bian, “Deep-learning-based Autonomous Navigation Approach for UAV Transmission Line Inspection”, in International Conference on Advanced Computational Intelligence (ICACI), 2018 10th IEEE.IEEE,2018, pp. 455-460.
[3] R. O. Duda, P. E. Hart, “Use of the Hough transformation to detect lines and curves in pictures,” Communications of the ACM, Vol.15, No.1, 1972, pp. 11-15.
[4] Tang Wen YANG, Hang YIN, Qiu Qi RUAN, “Overhead Power Line Detection from UAV Video Images”, in International Conference on Methatronics and Machine Vision in Practice(M2VIP), 2012 19th IEEE, 2012, 74-79.
[5] LeCun Y, Bengio Y, Hinton G, Deep learning[J], Nature, 2015.
[6] Long J, Shelhamer E, Darrell T, Fully convolutional networks for semantic segmentation, In: Proc, of the IEEE CVPR; 2015. p. 3431–40.
[7] He K, Gkioxari G, Dolla´r P, Girshick R, Mask R-CNN, arXiv preprint arXiv:170306870, 2017.
[8] K. He, G. Gkioxari, P. Dollár, R. Girshick, “Mask R-CNN” arXiv: 1703.06870v3, 2018.
[9] Ren, K. He, R. Girshick, and J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks. In NIPS, 2015.
[10] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu. Spatial transformer networks, In NIPS, 2015.
[11] J. Long, E. Shelhamer, and T. Darrell, Fully Convolutional Networks for semantic segmentation, In CVPR, 2015.
[12] Chenglong Wang, Guozhuang Chen, Qiuli Ma, Haoqun Fang, Xuan He, Zetao Chen, Peixuan Sun, Wire color recognition system software based on OpenCV, Electronic Measurement Technology, 2018.
[13] Congxin Zhang, Le Zhao, Xianpei Wang, Research on fast extraction algorithm of power line in complex ground object background, Engineering Journal of Wuhan University, 2018.
[14] Yongsheng Zhao, Haiqing Xu, Ligang Wu, Chong Liang, Research on Application of Foreign Object Recognition in Transmission Line Based on Hough Linear Transformation, Digital Technology and Application, 2017.
[15] Yuping Wang, Rearch and application on SVM-based Obstacle Recognition for High-voltage Power Line Inspection Robot, ChongqingUniversity, 2015.Conference Name:ACM Woodstock conference