# High-Voltage Power Transmission Tower Detection Based on Faster R-CNN and YOLO-V3

Hao Wang[1,2], Guodong Yang[1], En Li[1], Yunong Tian[1,2], Meng Zhao[3], Zize Liang[1]

1. The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation Chinese Academy of Sciences, Beijing 100190, China
E-mail: guodong.yang@ia.ac.cn

2. The University of Chinese Academy of Sciences, Beijing 100049, China
E-mail: wanghao2017@ia.ac.cn

3. Automation Department, North China Electric Power University, Hebei, Baoding, 071003, China
E-mail: meng.zhao@ncepu.edu.cn

**Abstract:** The power transmission mainly depends on overhead transmission infrastructures, such as towers and lines. Automatic inspection by robots or UAVs for the power transmission infrastructures is an essential way to ensure the safety of power transmission. Automatic detection and classification of the power towers is the prerequisite for automatic inspection. This paper compares two state-of-art deep learning methods to realize the high-voltage power transmission tower detection. We build the dedicated dataset of the power towers for multi-object detection, including data collection, preprocessing and annotation. After that, the models of YOLO-v3 and Faster R-CNN are used to solve multi-object detection on our dataset. The performances of the two models are evaluated under different indicators. It is verified that Faster R-CNN has a better detection performance in accuracy. However, the detection speed of YOLO-V3 model is faster and can be used in real-time detection.

**Key Words:** object detection, images acquisition, power tower detection, Faster R-CNN, YOLO-V3

## 1 Introduction

Nowadays, reliable electricity supply is related to the normal social economy and peoples' life. Restricted by geographical and environmental factors, the electric power production site is far away from the high consumption area. The overhead transmission line becomes the main means of power transmission. However, high-voltage power transmission tower as an important part of transmission lines, long-term exposure to the field environment, by the wind and snow erosion, coupled with continuous mechanical tension and material aging effects, there will be a variety of damage. Such as tower rust, cracking and other conditions. If these dangerous situations can't be dealt with in a timely manner, small damages may also expand and cause electric power transmission disruption, resulting in huge economic losses. So it is necessary for the overhead transmission line of the power tower to maintain and inspect regularly.

At present, the main ways of inspection of high-voltage transmission infrastructures are manual inspection, helicopter inspection and UAV inspection[1]. However, some of the high-voltage power transmission towers standing on top of the hill, it leads to workers need to reach the mountains, the efficiency of manual inspection is very low. Some lines pass through inaccessible wilderness areas, adding more security risks to the inspection. For particularly important power towers and lines, helicopters are used to cooperate with manual inspection. The helicopter is equipped with multiple camera sensors. The airborne platform includes a wide range of testing equipment such as thermal imaging and ultraviolet imaging. The technicians can check the test results at any time to determine the parts that need to be repaired, so that the workers can carry out subsequent repairs. The overall efficiency of the inspection is improved, but the driving technique is required to be close to the flight and the inspection operation cost is very high. It is difficult to apply it widely in the ordinary line inspection. Although the UAV inspection can reduce the inspection cost and safety risk, its inspection accuracy is low and the anti-interference is poor. The electric power industry is actively looking for automatic inspection solutions to solve the above problems[2]. At present, the detection and localization of the power towers is the prerequisite for automatic inspection. This has important guiding significance for the late maintenance of automatic inspection. Therefore, this paper uses the deep learning methods to realize the high-voltage power transmission tower detection and classification.

Most of the high-voltage power transmission towers are located in the mountains and forests. Uneven illumination conditions, complex background, camera shooting angle and distance and other factors can have a certain impact on the object detection. Whitworth et al. used template-matching method to extract features from power towers by establishing a visual tracking system to locate and track the power towers[3]. Sun et al. reconstructed the environment of transmission line through three-dimensional reconstruction, stereo vision and stereo matching, and found the position of the power tower in the image by classifying the power tower[4]. Golightly et al. use corner detection and matching algorithms to detect and locate power towers. At the same time, in the experiment of different illumination changes, the corner tracking algorithm can adapt to the contrast loss in a short time[5]. Cheng et al. proposed an approach to detect the power towers object from images. Using graph cut which is a developing graph based images segmentation technique, the prior knowledge and graph cut are combined to realize automatic image detection and localization[6]. Tragulunch

*This work was supported by the National Key R&D Program of China (2017YFC0806501, 2018YFB1307400) and the National Natural Science Foundation of China (61873267, U1713224).

et al. used the Canny edge detector to detect the edge features of the power towers and used the Hough transformation to detect the straight-line features of sub-images to locate the power towers[7]. Sampedro et al. applied a sliding window function to locate the power towers in the image and extract the HOG feature from image, which was passed as an input to the MLP classifier for training to separate the power tower from the background in the image[8]. Although different power tower detection and segmentation algorithms produced objective results, most of the results show only one type of power tower and simplify the complexity of the color, appearance and shape of the power tower. However, in actual situations, the color and materials used in the high-voltage power transmission towers are not the same, and the appearance and shape are also extremely diverse. With the continuous development of machine learning, the application of deep learning technology in the automatic inspection of transmission lines is becoming more and more extensive. Therefore, this paper uses supervised deep learning methods to detect and classify four different types of high-voltage power transmission towers. First of all, we collect the multi-angle, all-round images of the power towers as our dataset. Secondly, we use Matlab for data preprocessing and data augmentation, using Labelimg software for data labeling. Then we build two state-of-the-art object detection frameworks, Faster R-CNN and YOLO. We apply the constructed high-voltage power transmission tower dataset for training and testing, analyzing and evaluating the performance of the model through experimental results. Finally, we summarize and point out the future research directions.

## 2 Establishment of Our Dataset

### 2.1 Data Collecting

Most of the high-voltage transmission infrastructures are installed in areas such as mountains and in the wild where there are few people and the terrain is complicated. In order to ensure the normal operation of the transmission infrastructures, this not only brings the difficulty of inspection and maintenance, but also causes great difficulty in collecting the images. Most of the visual information is captured and collected by manned helicopters, but this increases the cost of inspections and operations. Based on the above situation, our images data is mainly composed of two parts: the ground shooting and the DJI M200 UAV equipped with the Pan-Tilt camera of Zemuse X4S. Fig.1 shows our data collection UAV system.

Fig. 1: UAV data acquisition system

Since power towers play an extremely important role in overhead transmission lines, different types of towers can be classified according to their use, shape, material and structure. In this paper, four common power towers with different shapes are collected as the research object. As shown in Fig.2, there are drum-shape tower, umbrella-shape tower, wineglass-shape tower and cathead-shape tower.

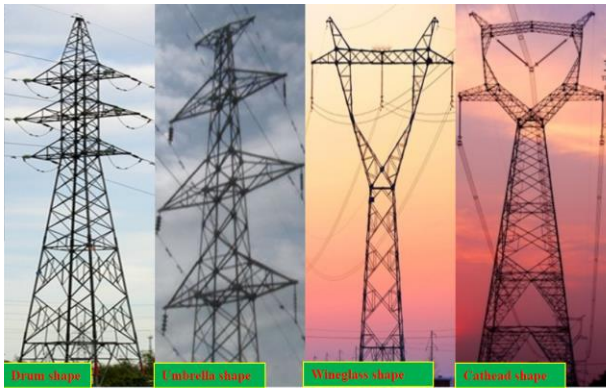


Fig. 2: Four different types of towers

### 2.2 Data preprocessing

#### 2.2.1 Data Augmentation

In the process of using the UAV to collect experimental data, since the angle and intensity of different sunlight are different during the day, whether the neural network model can detect and identify the image collected at different times of the day depends on the integrity of the dataset. At the same time, the UAV performs data acquisition during flight. Different shooting angles and shooting distances also affect the detection of the high-voltage power transmission tower. Therefore, in order to enhance the richness of the experimental dataset, we process the collected images from the aspects of brightness and angle to expand our dataset.

In this paper, we use Matlab to expand and enhance the dataset. Randomly select two values from 0.65 to 1.3 to adjust the brightness of original images, and save the two new results to the training set. This method can make up for the shortcoming of the neural network which is not robust to different illumination intensity caused by the concentration of image acquisition time. Then, we adopt three methods of rotating images 90, 180 degrees and mirroring to further augment the image dataset. The robustness of the neural network under different shooting angles can be enhanced by image data augmentation in angle aspect.

#### 2.2.2 Dataset Annotation

In order to meet the needs of subsequent experiments, images in the training set is made into Pascal VOC format. First, we rescale the long side of training set images to 500 pixels, adjusting the short side according to the same scaling ratio. The adjusted images are numbered uniformly.

Then we use Labelimg software to label our data manually. No labeling is done for positive samples whose pixel area is too small or very unclear to prevent neural network over-fitting caused by these samples. In the case of occlusion, the target whose occlusion area is greater than 70% and the target at the edge of the picture whose area is less than 30%

are not labeled. After each image is labeled, a corresponding XML file is generated, which contains the category of the object, the coordinates of the bounding box, and the size and depth of the image. Labeling images is a prerequisite for supervised learning. However, due to the time-consuming and laborious process of manual labeling, it may bring corresponding matching errors. This tool allows us to check the labeled images and related labels, and correct them in case there is a labeling mistake.

## 3 Model Construction

We hope that the model can detect and classify multiple objects at the same time. At present, the object detection networks with best performance and wide application include R-CNN, SPP-NET, Fast R-CNN, Faster R-CNN and YOLO. Therefore, we use two state-of-the-art deep learning networks, Faster R-CNN and YOLO-V3 in our experiment. Below, we briefly introduce two kinds of networks.

### 3.1 Faster R-CNN

The Faster R-CNN detection model is shown in Fig. 3. Faster R-CNN can be divided into the following four contents: The first content is the convolution layer. As an object detection method of convolutional neural networks, Faster R-CNN first uses a set of basic convolutional layers, relu layer and pooling layer to extract feature map in an input image. The feature map is shared for subsequent RPN layers and fully connected layers. The second content is the RPN (Region Proposal Networks). The RPN is used to generate the region proposals. This layer uses Softmax to determine whether the anchors belong to the foreground or the background, and then uses the bounding box regression to rectify the anchors to obtain accurate objects. The third content is the RoI (Region of Interest) Pooling layer. This layer collects the feature map and proposals, synthesizing the information and extracting the proposal feature maps. Then the proposal feature maps are sent to the subsequent fully connected layer to determine the classes of object. The last content is classification. The proposal feature maps are used to calculate the classes of proposals, and the final precise position of the target box is obtained by the bounding box regression.

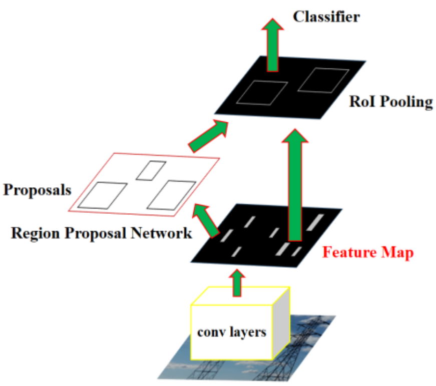


Fig. 3: The structure of Faster R-CNN

This model uses alternating training. It gets initial parameters from pre-trained VGG16. After obtaining parameters, we train RPN firstly and then send the result to Fast R-CNN. We adjust the weights of convolution layer. Finally, RPN get data from convolution layer.

### 3.2 YOLO

One feature of the deep learning approach is the end-to-end detection. Different from other methods of object detection, dividing the tasks of object detection into multiple processes such as region-proposal prediction and classified prediction. YOLO integrates region-proposal prediction and classified prediction into a neural network model. It realizes fast object detection and recognition with high accuracy. This method is more suitable for field application environments. In Faster R-CNN, although RPN and Fast R-CNN share convolution layer, in the process of training model, it needs to train RPN and Fast R-CNN repeatedly.

Different from "look twice" process of the series of R-CNN, YOLO only has to look once. R-CNN separates the detecting process into two: object detection, which is a classification problem, and the position of bounding box, which is a regression problem. In contrast, YOLO combines them into a regression problem.

The YOLO detection model is shown in Fig. 4. The network divides the input image of training set which has the object of high-voltage power transmission tower into S×S grids. If the center of the object ground truth falls in a grid, the grid is responsible for detecting the object. Each grid predicts B bounding boxes and their confidence scores, as well as C class conditional probabilities. Confidence score is defined as follows.

$$Confidence = Pr(Object) * IOU_{pred}^{truth}\ ,\ Pr(Object) \in (0,1) \quad (1)$$

When the target is in the grid, $Pr(Object) = 1$, otherwise 0. $IOU_{pred}^{truth}$ is used to denote the coincidence between the real ground truth and the predicted bounding box. The confidence reflects whether the grid contains objects and the accuracy of the predicted bounding box when it contains objects. When multiple bounding boxes detect the same target, YOLO uses the non-maximum suppression (NMS) method to select the best bounding box as the final one.

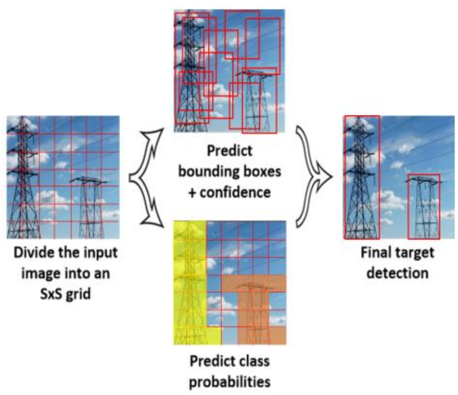


Fig. 4: The method of YOLO detection

### 3.3 Model comparison

Faster R-CNN and YOLO are good object detection methods. Faster R-CNN constructively proposed the RPN structure. After the convolutional neural network, RPN was added as the branch network to realize the extraction of the anchor box and merged into the deep network. Faster R-CNN proposed RPN as a regional selection network, which realizes the function of the neural network to select the detection region. However, it still stays in the thought of "two-stage", and still reflects the process of region proposal concretely.

For YOLO, through the continuous improvement of defects, the current YOLO-V3[11] network, which is evolved from YOLO-V1[9] and YOLO-V2[10] networks. YOLO-V1 directly regresses the position and class of the bounding box in the output layer, making the detection speed very fast. However, the accuracy of identifying the position of the object is not high. Although each grid can predict B bounding boxes, only the highest bounding box of IOU is selected as the output of object detection, that is, each grid predicts at most one object. When each grid contains multiple objects, only one of them can be detected. In order to improve the accuracy of object localization, YOLO-V2 introduces the idea of "anchor box" in Faster R-CNN, and uses k-means clustering algorithm to generate suitable prior bounding boxes. At the same time, the IOU is involved in distance calculations, resulting in better IOU values through these anchor boxes. In addition, YOLO-V2 improves the design of the network structure and uses convolution layer to replace the fully connected layer of YOLO-V1 in the output layer. YOLO-V3 uses multi-scale prediction to detect the final object based on YOLO-V2, and the network structure is more complex and the detection performance is stronger. Therefore, this paper uses YOLO-V3 to detect and classify high-voltage power transmission towers.

## 4 Experiment and Results

This paper uses the two best frameworks, Faster R-CNN and YOLO-V3, to implement the detection and classification of power towers. The detection models are trained and tested on an NVIDIA Tesla V100 server.

### 4.1 Indicators of Evaluation model

The related indicators for evaluating the effectiveness of the two models are as follows:

***A. Loss Function***

Loss function is one criterion for evaluating the performance of a model. In Faster R-CNN, the loss function for an image is defined as:

$$L\left(\{p_i\},\{t_i\}\right)=\frac{1}{N_{cls}}\sum_i L_{cls}\left(p_i,p_i^*\right)+\lambda\frac{1}{N_{cls}}\sum_i p_i^* L_{reg}\left(t_i,t_i^*\right) \tag{2}$$

Where $i$ is the index of an anchor in a mini-batch and $p_i$ is the predicted probability of anchor $i$ being an object. The ground-truth label $p_i^*$ is 1 if the anchor is positive, and is 0 if the anchor is negative. $t_i$ is a vector representing the 4 parameterized coordinates of the predicted bounding box, and $t_i^*$ is that of the ground-truth box associated with a positive anchor. The classification loss $L_{cls}$ is log loss over two classes (object vs. not object). For regression loss, we use:

$$L_{reg}\left(t_i,t_i^*\right)=\sum_{u\in\{x,y,w,h\}} smooth_{L_1}\left(t_i-t_i^*\right) \tag{3}$$

In which

$$smooth_{L_1}\left(x\right)=\begin{cases}0.5x^2 & if \quad |x|<1\\ |x|-0.5 & otherwise\end{cases} \tag{4}$$

The term $p_i^* L_{reg}$ means the regression loss is activated only for positive anchors $p_i^*=1$ and is disabled otherwise $p_i^*=0$.

The loss function of YOLO is defined as follows:

$$\begin{aligned}Error_{coord} &= \lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{obj}\left[\left(x_i-\hat{x}_i\right)^2+\left(y_i-\hat{y}_i\right)^2\right]\\ &+\lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{obj}\left[\left(\sqrt{w_i}-\sqrt{\hat{w}_i}\right)^2+\left(\sqrt{h_i}-\sqrt{\hat{h}_i}\right)^2\right]\\ &+\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{obj}\left(C_i-\hat{C}_i\right)^2\\ &+\lambda_{noobj}\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{noobj}\left(C_i-\hat{C}_i\right)^2\\ &+\sum_{i=0}^{S^2}1_i^{obj}\sum_{c\in classes}\left(p_i(c)-\hat{p}_i(c)\right)^2\end{aligned} \tag{5}$$

Where $\lambda_{coord}$ is the weight of the coordinate error, $s^2$ is the number of grids in the input image, and B is the number of bounding boxes generated by each grid. Referring to the original parameters in the YOLO-V3 model, $\lambda_{coord}=5$, $S=7$, B=9 were selected in this study. $1_{ij}^{obj}=1$ denotes that the object falls into the $j_{th}$ bounding box in grid $i$, otherwise $1_{ij}^{obj}=0$. $\left(\hat{x}_i,\hat{y}_i,\hat{w}_i,\hat{h}_i\right)$ are values of the center coordinate, height, and width of the predicted bounding box. $\left(x_i,y_i,w_i,h_i\right)$ are true values. The parameter $\lambda_{noobj}$ is the weight of the IoU error. Referring to the original parameters of the YOLO-V3 model, $\lambda_{noobj}=0.5$ was selected in this paper. $\hat{C}_i$ is the predicted confidence, and $C_i$ is the true confidence. The $c$ refers to the class to which the detected target belongs. $p_i\left(c\right)$ refers to the true probability that the object belonging to class $c$ is in grid $i$. $\hat{p}_i\left(c\right)$ is the predicted value. The $Error_{cls}$ for grid $i$ is the sum of classification errors for all the objects in the grid.

***B. Intersection over Union(IoU)***

IoU is a standard for defining the detection accuracy of target objects. IoU evaluates the performance of the model by calculating the overlap ratio between the predicted bounding box and the true bounding box as follows:

$$IoU=\frac{S_{overlap}}{S_{union}} \tag{6}$$

where $S_{overlap}$ is the area of intersection of the predicted bounding box and the true bounding box. $S_{union}$ is the area of the union of the two bounding boxes.

***C. Detection Time***

The average detection times for two deep learning models were compared in this paper, and the real-time performance of these models was analyzed.

***D. Confusion Matrix***

For the four-classification problem of this paper, we construct the confusion matrix to evaluate the accuracy of the classification of models according to the combinations of the actual class and predicted class.

## 4.2 Analysis of Experimental Results

In order to verify the feasibility of the two models used in this paper, the loss curves of YOLO-V3 and Faster R-CNN are shown in Fig. 5 in the training stage.

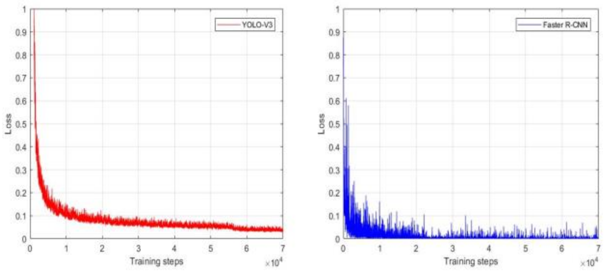


Fig. 5: The Loss curves of two models

In the training stage, we can see that the loss of the two models are convergent. The loss curve of the YOLO-V3 model begins to flatten after 5000 training steps, and the loss no longer decreases after 55000 steps. The final loss of YOLO-V3 is around 0.017. Although there are some oscillations in the convergence process of the loss of the Faster R-CNN model, it begins to flatten at around 3000 steps, and the convergence speed is very fast. The final loss is around 0.0031. Therefore, the two models used in this paper can be used to train and detect the images of high-voltage power transmission tower.

The IoU and average detection time of the two models are shown in Table 1.

The detection results of the two models for different types of towers are shown in the figures. The results of the Faster R-CNN model are shown in Fig. 6, and the results of the YOLO-V3 model are shown in Fig. 7.

The confusion matrixes of the two models in the classification stage are shown in Table 2 and Table 3.

In the detection indicators, we can see that the IoU of the Faster R-CNN model is 0.882, which is higher than 0.874 of the YOLO-V3 model. This indicates that Faster R-CNN is better than YOLO-V3 in terms of detection accuracy. However, the average detection speed of the YOLO-V3 model is faster than that of the Faster R-CNN model, so it has a greater advantage in the real-time detection.

For the confusion matrix, we can see that the Faster R-CNN model has a better classification result with the same test samples. Specifically, the Faster R-CNN model has 100% accuracy for the drum-shape tower and the umbrella-shape tower. This is relatively easy understanding, for the features of these two types are very obvious and distinguishable. For the wineglass-shape tower and the cathead-shape tower, the characteristics of these structures are more complicated. In addition, the changes of the environment condition, shooting parameters and the clarity of the test samples will all affect the classification accuracy.

Table 1: Detection Indicators of Two Models

| Detection indicators / Model Types | Average IoU | Average time (s) |
|---|---|---|
| Faster R-CNN | 0.882 | 2.136 |
| YOLO-V3 | 0.874 | 0.0215 |

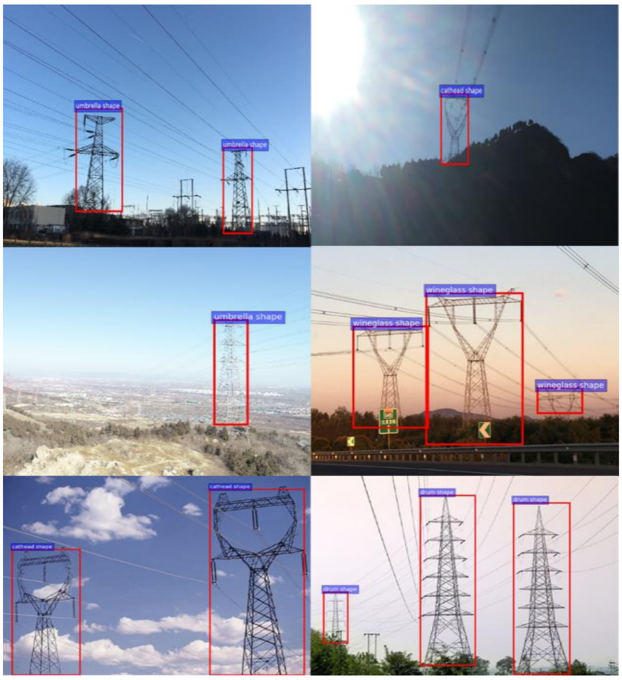


Fig. 6: Tower detection results of the Faster R-CNN model

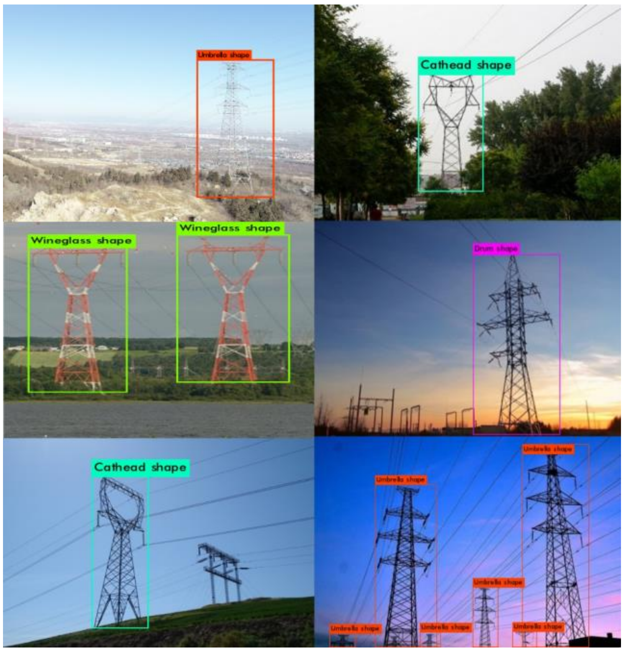


Fig. 7: Tower detection results of the YOLO-V3 model

Despite of those adverse conditions, the classification results are quite satisfactory, with above 94% accuracy for Faster R-CNN model. From the confusion matrix of the YOLO-V3 model classification results, it can be seen that there is no better classification accuracy, which is caused by the difference in feature detection method between the two models. YOLO model combines the classification and location of the target into a regression problem. Although the detection speed is fast, the detection accuracy is correspondingly reduced.

From the above comparisons, we can see that both Faster R-CNN and YOLO-V3 can be used to detect and classify the high-voltage power transmission tower. Faster R-CNN has better performance in detection accuracy, and YOLO-V3 is better in real-time performance.

Table 2: Confusion matrix of Faster R-CNN model

| Actual class / Predicted class | Drum shape | Umbrella shape | Wineglass shape | Cathead shape |
|---|---|---|---|---|
| Drum shape(%) | 100 | 0 | 0 | 0 |
| Umbrella shape(%) | 0 | 100 | 0 | 0 |
| Wineglass shape(%) | 0 | 0 | 94 | 4 |
| Cathead shape(%) | 0 | 0 | 6 | 96 |

Table 3: Confusion matrix of YOLO-V3 model

| Actual class / Predicted class | Drum shape | Umbrella shape | Wineglass shape | Cathead shape |
|---|---|---|---|---|
| Drum shape(%) | 96 | 2 | 2 | 2 |
| Umbrella shape(%) | 4 | 98 | 2 | 0 |
| Wineglass shape(%) | 0 | 0 | 88 | 6 |
| Cathead shape(%) | 0 | 0 | 8 | 92 |

## 5 Conclusion

For the automatic inspection of high-voltage transmission line, the detection and classification of high-voltage power transmission towers are the preliminary problems. We use two deep learning methods, Faster R-CNN and YOLO-V3, to verify the performances of power tower detection with different types in high-voltage power transmission environment. In this paper, the images of various forms of power towers are collected in the field environment of high-voltage power transmission firstly. All the images are enhanced by transforming the angles and brightness in the training set. Then, the two deep learning models are used to solve the multi-object detection and classification problems of the power tower. From the experiment results, we can see that the loss of the two models are convergent in the training stage. The final loss of YOLO-V3 is around 0.017, and the final loss of Faster R-CNN is around 0.0031. The average IoU of the Faster R-CNN is 0.882, which has better performance for localization of the high-voltage power tower. In addition, for the four-classification problem of this paper, we gave the confusion matrix of the two models. In the test set, Faster R-CNN has a better performance in detection accuracy. However, the detection speed of YOLO-V3 model is faster, and the average detection time is 0.0215s, which can be realized in real-time detection.

In the future, the detection method of power tower verified in this paper can be used to realize the real-time localization and tracking of automatic inspection by UAV. At the same time, we will also develop and expand the defect detection and analysis system to realize the automatic inspection.

## References


[1] J. Katrasnik, F. Pernus, and B. Likar, A survey of mobile robots for distribution power line inspection, *IEEE Transactions on Power Delivery*, 25(1): 485-493, 2009.

[2] A. Pagnano, M. Höpf, and R. Teti, A roadmap for automated power line inspection maintenance and repair, *Procedia CIRP*, 12: 234-239, 2013.

[3] C. C. Whitworth, A. W. G. Duller, D. I. Jones and G. K. Earp, Aerial video inspection of overhead power lines, in *Power Engineering Journal*, 15(1):25-32, 2001.

[4] C. Sun, R. Jones, H. Talbot, X. Wu, K. Cheong, R. Beare, M. Buckley, and M. Berman, Measuring the distance of vegetation from power lines using stereo vision, *ISPRS journal of photogrammetry and remote sensing*, 60(4):269–283, 2006.

[5] I. Golightly and D. Jones, Corner detection and matching for visual tracking during power line inspection, *Image and Vision Computing*, 21(9):827–840, 2003.

[6] W. Cheng and Z. Song, *Power Pole Detection Based on Graph Cut*, *2008 Congress on Image and Signal Processing*, Sanya, Hainan, 2008:720-724.

[7] P. Tragulnuch, T. Chanvimaluang, T. Kasetkasem, S. Ingprasert and T. Isshiki, High Voltage Transmission Tower Detection and Tracking in Aerial Video Sequence using Object-Based Image Classification, *2018 International Conference on Embedded Systems and Intelligent Technology & International Conference on Information and Communication Technology for Embedded Systems (ICESIT-ICICTES)*, Khon Kaen, 2018:1-4.

[8] C. Sampedro, C. Martinez, A. Chauhan and P. Campoy, A supervised approach to electric tower detection and classification for power line inspection, *2014 International Joint Conference on Neural Networks (IJCNN)*, Beijing, 2014:1970-1977.

[9] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE conference on Computer Vision and Pattern Recognition*, 2016:779-788.

[10] J. Redmon and A. Farhadi. YOLO9000: Better, Faster, Stronger. *2017 IEEE conference on Computer Vision and Pattern Recognition*, 2017:6517-6525.

[11] J. Redmon and A. Farhadi. YOLOv3: An Incremental Improvement. *2018 IEEE conference on Computer Vision and Pattern Recognition*, (p. arXiv:1804.02767).