Transmission Line Component Inspection Method Based on Deep Learning under Visual Navigation

Liu Yue1,2*, Yang Yuechen1,2, Sun Xiaobin3, Li Jianxiang1,2, Zhang Huadong3

1 State Grid Shandong Electric Power Research Institute, Jinan, Shandong, 250000, China
2 Shandong Luneng Intelligence Technology Co., Ltd. Artificial Intelligence on Electric Power System, State Grid Corporation Joint Laboratory, Jinan, Shandong, 250020, China
3 State Grid Shandong Electric Power Company, Jinan, Shandong, 25000, China
*Corresponding author’s e-mail: liuyue2046@126.com

Abstract. The efficiency of the operation and maintenance for transmission lines can be greatly enhanced with the help of UAV. However, the images of the transmission line component collected by UAV could be spoiled due to the incorrect operations of the controllers, which causes the defects of the transmission line component couldn’t be diagnosed accurately. A visual inspection algorithm for the UAV based on the sensing of the tower components is proposed which uses a parallel network structure and can be applied on embedded devices, which based on Faster R-CNN target recognition framework and the ZFNet. Thus, the classification and posture of the component can be acquired. The UAV’s pose and position can be adjusted by the distance sensing model and components’ posture. Results show that the maximum mAP (mean Average Precision) of component recognition and posture perception are 0.95 and 0.91, respectively. Additionally, the posture of the UAV can be adjusted effectively and optimum images of the component can be obtained.

1. Introduction
With the development of unmanned aerial vehicle (UAV) technology, transmission lines inspection with UAV becomes more and more widely used. It has greatly improved the maintenance efficiency. The inspector should also operate the camera to capture tower component images. And it’s difficult to obtain the standard and best inspection images in a right shooting angle. Besides, the transmission line infrastructure should be inspected multiple times, which cannot guarantee the quality of inspection image. Therefore, this paper proposes a transmission line component pose inspection method to navigate the UAV and achieve the intelligent inspection.

The research of pose recognition mainly consists of two solutions: one is to use two-dimensional model and the other is to use the 3D information to solve the problem. Some researchers proposed the pose detector, which uses variable features, such as HOG[1] and deformable models[2], which based on two-dimensional image information. Redondo et al. proposed a simultaneous object recognition and pose estimation method based on Hough Forest, the recognition performance is not as good as the method which based on DPM (Deformable Parts Model)[3]. The latest development of the pose estimation model is to use the 3D CAD model[4]. Bojan et al. introduced the 3D extension of the DPM model, in which the component appearance and spatial deformation is represented in 3D. Zia et
al. used off-the-shelf methods to locate targets roughly, and then used an annotated 3D CAD model to estimate continuous poses[5]. However, the two kinds of model treat the pose recognition problem as a multi-classification problem, using a pre-trained model to classify the component. In addition, the cost of constructing a 3D CAD model of tower components is very huge.

With the huge success of deep learning method in the field of large-scale target recognition, some target detection methods based on CNN are gradually proposed[6]. H Kai et al. proposed the Faster R-CNN structure, which uses the RPN network to generate proposed area, decrease the time complexity by the traditional area proposal method[7]. However, the above target recognition system generally uses ZFNet[8], VGG16[9], ResNet101[10] and other CNN to extract features. When the network is growing deeper, the model contains too many parameters and high computational complexity. The complex model is difficult to integrate effectively in our embedded system. Because it requires a lot of calculations time, and it is difficult to integrate effectively in the embedded system. Recently, J Redmon et al. proposed the YOLOv3 [11] network structure. It has a 106-layer full convolution architecture, the structure is more complicated. The trained model is more complex and the recognition speed has no significant improvement. YOLOv3 has the best average recognition result in the YOLO series, and the recognition speed is faster than Faster R-CNN. But the mAP of the experimental results is not as good as the Faster R-CNN network structure. Besides, it is difficult to meet the practical application requirements too.

In order to take the standard and the best inspection image of the components, this paper proposes a tower components’ posture recognition algorithm, which inspects tower pose and navigates UAV with the visual information extracted from captured image. Besides, the visual perception control model of the UAV is established to adjust the UAV’s pose.

2. Transmission line component pose perception

2.1. Analysis of the structure of the tower components

The tower components’ image captured by the UAV is shown in figure 1. From the analysis of the transmission line structure, there are two main types of towers: straight tower and tension-resistant tower. When the transmission line passes through a linear tower, it is connected by an insulator to form a tensile force. So that the line presents a fold line (or an acute angle connection), which forming a connected region. As shown in figure 2, there are parallel lines and drain lines in the tower, the drain lines are connected with the parallel lines at both ends. When the line pass through the tension-resistant tower, the two ends of the line are connected by a downwardly protruding drain line. Therefore, the parallel straight line group and the parallel curve group are connected. The angle between the two is shown in figure 3.

![Figure 1](image1.png)
Figure 1. Several posture of the tower captured by the actual drone inspection.

![Figure 2](image2.png)
Figure 2. Line diagram of the wire passing through the straight tower.

![Figure 3](image3.png)
Figure 3. Relationship between the drain line and the parallel line.
These connected areas are the key parts of tower inspection points of transmission line, including equalizing rings, triangular plates, insulators etc.. As shown in (a) to (f) of figure 1, tower components’ posture are divided into three types: front side, left side and right side. In conclusion, transmission line’s head and body are surround by them. The metal fixture components are directly or indirectly connected to the insulator, and they are gathered near the insulator too. Thus, the other connected components’ type and posture can be inferred by recognising the position and posture of the insulator. The connecting components are symmetrically distributed on the left and right sides of the tower, where the insulator acts as an insulation function at the junction of tower and line. The equalizing ring is between insulator and transmission line, and they have an up-down or left-right positional relationship with the insulator. It can be seen that, there is also a spatial context relationship between upper and lower sides of all connecting components. Based on the context relationship, after the posture of a certain component is recognised, the posture of the other connecting components can be estimated.

2.2. Visual tower component sensing method
To capture the standard and best component images, the transmission line components and their pose will be inspected during the inspection phase. So that the position and posture of the UAV can be adjusted according to the posture of the tower components. Based on the Faster R-CNN network, the classification method of the tower components is proposed. As is shown in figure 4, component recognition and posture perception are realized at the same time. Among the parallel network structure, the first network represents a classification network, the second network represents a posture recognition network. Classification network is responsible for classifying several types of transmission line components, posture recognition network is responsible for recognise the pose of the tower. Both the two network are based on the Faster R-CNN network framework.

Combined with the category of the tower components and the actual posture, the classification data set and the pose data set are respectively constructed. Based on Caffe[12], a deep learning framework, the classification model and the pose recognition model were trained separately. Three different feature extraction networks, such as ZFNet, VGG16 and ResNet101, were used to propose component features for comparative experiments. The experiment is divided into training phase and testing phase, in which the training phase uses the constructed data set respectively, to train corresponding classification model and pose perception model. In the test phase, the trained component type model and the component pose recognition model obtained are loaded firstly. Then, the collected image samples are respectively input into the component type classification model and the pose perception model. Finally, the convolutional neural network features are extracted and then the type and the pose of the component are outputted.

![Network Diagram](image)

**Network 1**
- Data set
- Data preprocessing
- CNN layer
- Fully connected layer
- Softmax classification
- Target recognition result
- Pose and category

**Network 2**
- Data set
- Data preprocessing
- CNN layer
- Fully connected layer
- Softmax classification
- Pose recognition result

Figure 4. Component Identification Gesture Detection Parallel Network Structure.

2.3. Embedded tower component recognition system
To control the UAV to reach the nearby position of the tower component, it is necessary to determine the geographical position of the component to be recognised. A position navigation system is designed
and implemented, which based on the perceived posture of the tower components, by taking a global shot of the whole and then focusing on the local target.

The development process shown in figure 5 is constructed with multi-threading technology, where the thread A is responsible for image acquisition, and the captured image is acquired from the on-board camera in real time. The thread B is used to recognise the posture of the tower component. The serial port communication is established between the NVIDIA TX2 device and the UAV flight control system. The vector of the recognised tower components posture is converted into a control vector of UAV, which used to transfer to the UAV’s control system. By establishing the relationship between the posture of the tower component and the direction of the UAV movement, the feedback system is established.

![Figure 5. Parallel development flow chart.](image)

3. Visual based tower component sensing and UAV’s pose control

The distance sensing model based on monocular vision model[13], which sense the tower component from far position to near position as shown in figure 6. During the inspection, the insulator is around the tower area, which can be recognised very well. Therefore, the insulator is selected as our inspection targets. Assume that the UAV flies forward in a straight line from far position to near position. As shown in figure 6, point A and point B are far and near point, respectively. \( y_a \) and \( y_b \) are the height of the insulator in the pixel plane. \( l_a \) and \( l_b \) is the image in the respective camera planes at the A and B positions of the tower component \( L \). \( f \) is the camera focal length.

![Figure 6. Visual distance perception model based from far position to near position.](image)

Whereas \( \Delta y_1 \) and \( \Delta z_1 \) are the width and height of a given object measured in the pixel plane. \( \Delta y_a \) pixels and \( \Delta y_b \) are the heights of the insulators in the pixel plane. The distance \( d_{AB} \) is the distance between point A and B. It can be obtained by integrating the speed. The actual width and height is \( Y \) mm and \( Z \) mm. The geometric relationship of the monocular vision model is known as:

\[
\frac{f}{\Delta y_a} = \frac{d_{BO} + d_{AB}}{Y}, \quad \frac{f}{\Delta y_b} = \frac{d_{BO}}{Y}
\]

(1)

According to the change of image ratio \( S \) between \( \Delta y_a \) and \( \Delta y_b \), the distance between UAV and components can be calculated as formulas (2)

\[
\frac{\Delta y_a}{\Delta y_b} = S = \frac{d_{BO}}{d_{BO} + d_{AB}}
\]

(2)

Then the distance \( d = d_{BO} \) can be expressed as:

\[
d_{BO} = \frac{d_{AB}}{1 - S}
\]

(3)

The model established above is an ideal situation. Actually, the posture of the UAV which relates to the insulator can’t be guaranteed. Through the analysis of the insulators position on a large number of towers, the insulators have only three pose on the tower: horizontal, vertical and inclined 45°. And there are three main types of posture between the actual insulator and the UAV: front side, left side, right side. The distance of three pose is shown in in figure 7.
Figure 7. The UAV position and distance to the tower. (a) front side, (b) left side, (c) right side.

The distance between the UAV and the tower component can be estimate easily according to the insulator. Then, UAV need to change pose to get standard picture. According to the working principle of the IMU [14], the current UAV’s pose, the perceived posture, the distance between the tower components and UAV are combined together to ensure the UAV pose. The pose adjustment method is shown in figure 8.

Figure 8. UAV pose correction, (a) right side, (b) front side, (c) left side.

The specific adjustment method is:

1) When the posture of the identified tower component is the left side, it is necessary to rotate the yaw angle about 45° clockwise along the X-axis and fly a distance of $\sqrt{2}/2d$ in the Y-axis direction along the new yaw.

2) When the posture of the identified tower component is the right side, it is necessary to rotate the yaw angle about 45° counterclockwisely along the X-axis and fly a distance of $\sqrt{2}/2d$ in the Y-axis direction along the new yaw.

4. Experiment and analysis

4.1. Data set preparation

There is no public data set in the transmission line inspection task at present. The proposed method uses images of different transmission line components collected by the laboratory, mainly from the taken pictures during the inspection of the UAV. As shown in figure 9, triangular plate, insulator and equalizing ring are selected as the training model for classification model, and three posture of these tower components are selected as the training samples for posture perception model. Each component has three types of posture. Each category has 1000 sheets high-resolution images, a total of 9000 original samples. And the ratio of the training set to the test set is 9:1.

4.2. Classification and identification experiments and analysis

The experiment uses the transmission power classification data set, using insulators, circular equalizing rings, and triangular plates for three types of transmission line components. Based on the Faster R-CNN framework, the three convolution networks of ZFNet, VGG16 and ResNet101 are used as feature extraction network.

| Net      | Recall rate | Correct rate | Leak recognition rate |
|----------|-------------|--------------|-----------------------|
| Component type | insulator | equalizing rings | mAP | insulator | equalizing rings | mAP | insulator | equalizing rings | mAP |
| ZFNet    | 0.833       | 0.746        | 0.790             | 0.983            | 0.971        | 0.977 | 0.153    | 0.232            | 0.123 |
| VGG16    | 0.877       | 0.779        | 0.828             | 0.991            | 0.995        | 0.993 | 0.115    | 0.192            | 0.154 |
Table 2. Time required for three models to detect a single image on NVIDIA TX2 embedded platform.

| Feature extract network | ZFNet5 | VGG16 | ResNet101 |
|-------------------------|--------|-------|-----------|
| Run Time in NVIDIA TX2  | 0.27s  | 1.9s  | 8.4s      |

As shown in table 1, the inspected results have some difference based on three convolution feature extraction networks. To a certain extent, the deeper the network, the better the recognition. The correct rate and recall rate of mean mAP nearly 0.97 and 0.79 of the ZFNet closely to the VGG16 and ResNet101.

4.3. Pose perception experiment and analysis

After the proposed network is trained by the server, the obtained model is run on NVIDIA TX2 device to recognize the posture of the tower. The statistical results in these types posture are shown in table 3 and table 4.

Table 3. Pose classification correct rate of insulators under different characteristic networks.

| Network Type | Front | Left | Right | mAP (pose) | Time complexity (s/frame) |
|--------------|-------|------|-------|------------|--------------------------|
| ZFNet        | 0.893 | 0.896| 0.964 | 0.918      | 0.24                     |
| VGG16        | 0.901 | 0.912| 0.931 | 0.914      | 1.7                      |
| ResNet101    | 0.889 | 0.895| 0.929 | 0.904      | 7.9                      |

Table 4. Pose classification correct rate of equalizing rings under different feature networks.

| Network Type | Front | Left | Right | mAP (pose) | Time complexity (s/frame) |
|--------------|-------|------|-------|------------|--------------------------|
| ZFNet        | 0.813 | 0.836| 0.824 | 0.824      | 0.25                     |
| VGG16        | 0.801 | 0.812| 0.831 | 0.814      | 3.1                      |
| ResNet101    | 0.809 | 0.855| 0.829 | 0.831      | 9.2                      |

According to the comprehensive table 3 and table 4, the accuracy of the recognition of tower components under different feature extraction networks is closely. According to Table 2, the processing speed of ZFNet is about 1.8 times and 12.9 times higher than VGG16 and ResNet101, ZFNet is suitable for running on embedded system.

5. Conclusion

The method is based on the Faster R-CNN architecture and the shallow feature extraction network. And it is applied to the embedded system to realize the classification task and pose recognition task. Using ZFNet as the feature extraction network to ensure the recognition speed is basically useful in application. Compared to single thread, this parallel recognition method reduces the processing time of the image recognition. It can be approximated in real time on the constructed UAV visual navigation tower inspection equipment. Based on the sensed component posture and the proposed visual perceptual feedback model, the position and pose of UAV is adjusted immediately. In the next stage, how to combine the more component’s type recognition and posture recognition in the same network will be researched.

Acknowledgements

This work was financially supported by the technology project of State Grid Shandong Electric Power Company (Application of remote airborne emergency UAV system in power industry, Grant No. 2018A-034), and National Key R&D Program (Key Technology Research and Application Demonstration of Unmanned Emergency Rescue Equipment, Grand No. 2017YFC08067). Thanks very much.

References

[1] Ozuysal M, Lepetit V, Fua P, et al. Pose estimation for category specific multiview object
[2] Gu C, Ren X. Discriminative mixture-of-templates for viewpoint classification[C]. European conference on computer vision, 2010: 408-421.

[3] Redondo-Cabrera C, López-Sastre R, Tuytelaars T. All together now: Simultaneous object detection and continuous pose estimation using a hough forest with probabilistic locally enhanced voting[C]. BMVC, 2014, 5: 6.

[4] Pepik B, Gehler P, Stark M, et al. 3D 2 PM–3D deformable part models[C]. European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2012: 356-370.

[5] Zia M Z, Stark M, Schiele B, et al. Detailed 3D Representations for Object Recognition and Modeling[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(11): 2608-2623.

[6] He K, Zhang X, Ren S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[C]. European conference on computer vision, 2014: 346-361.

[7] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(6): 1137-1149.

[8] Zeiler M D, Fergus R. Visualizing and Understanding Convolutional Networks[J]. European conference on computer vision, 2014: 818-833.

[9] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[10] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[11] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[12] Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional architecture for fast feature embedding[C]. Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014: 675-678.

[13] Bogue R W. Machine Vision Theory, Algorithms, Practicalities[J]. Assembly Automation, 2012, 25(3):58.

[14] Ru Xiangyu, Jin Chao, Pan Chengfeng, Xu Chao. Application of Monocular Vision Inertial Fusion Method in Position Estimation of Unmanned Aerial Vehicles[J]. Control and Information Technology, 2018(06): 50-58.