Research on the Application of YOLOv4 in Power Inspection

Xingda Sun*, He Hao, Yuan Liu, Yuanyuan Zhao, Yimeng Wang and Ye Du

Beijing Guodian Tong Network Technology Co., Ltd., Block A, Dongxu International Center, Fengtai District, Beijing.
Email: sunxingda100@163.com*

Abstract. In order to meet the business needs of power inspection. The requirement is to identify the target quickly and in batches. In this paper, yolov4 convolution neural network is used as the technical to realize the target detection process of power inspection photos. Firstly, the training data set of power patrol inspection is accurately labeled by labeling. Then it is trained by the deep learning framework of Darknet. And the results are satisfactory. The test results show that the Precision is 0.92 and the Recall is 0.904 after training by the self-designed training data set of electric power inspection target detection. This detection effect can meet part of the requirements of target detection in power inspection. However, there are two problems in this test: the lack of data annotation set in power inspection and the serious overlap of different objects in some photos of the data set. The author of this paper hopes the relevant researchers to correct and supplement.

1. Introduction
In recent years, the increasing demand for electricity in all walks of life has accelerated the intelligent transformation and upgrading of the power industry. Among them, the performance of inspections is the application of intelligent equipment such as drones. According to the latest data, the scale of my country's UAV power inspection market in 2019 is close to 3 billion yuan, and the development is very gratifying. In the future, with the accelerated implementation of new infrastructure and the further popularization of 5G networks and artificial intelligence, the industry is quite optimistic about the future of electric power inspection drones.

From the application of power inspection scenarios, the advantages of drones are very obvious.

On the one hand, drones can replace most of the manual inspection work. It not only has the convenience of manual inspection, but also can overcome the shortcomings of manual inspection, such as being vulnerable to terrain restrictions, blind spots in the inspection range, low inspection efficiency, insufficient inspection time and durability, and so on.

On the other hand, the data collected by drone inspections has also exploded compared with traditional power inspections. In the era of big data, the collection and analysis of data will greatly promote industry management and decision-making, thereby promoting the precise and intelligent development of the industry. As an important platform, UAVs can integrate various technologies and devices such as cameras and sensors. These hardware devices can not only help obtain image and video information, but also perform high-speed transmission, analysis and processing, thus becoming an important data basis. Therefore, how to effectively process the collected data becomes a top priority. This paper is based on the YOLOv4 convolutional neural network, and applies it to the power inspection business scenario to improve the accuracy and speed of the power inspection result judgment.
2. YOLOv4 Target Detection

2.1. Implementation Process of YOLOv4

Inputting an image, three initial feature layers will be extracted from the CSP arknet53 of the backbone feature extraction network for target detection. The three initial feature layers are located in different positions of csparknet53, and their shapes are (608,608,32), (304,304,64), (152,152,128). These three feature layers are used to detect small, medium and large targets respectively.

After three initial feature layers are extracted from CSP arknet53, three effective feature layers can be obtained after certain processing. The shapes were (76,76,256), (38,38,512), (19,19,1024). Compared with yolov3, yolov4 mixes spp structure into convolution of the last feature layer of cspdarknet53. The last feature of netv253 is on the layer of darkpv253_BN. After leaky convolution, four different scales of maximum pooling are used for processing. The sizes of the pool nuclei with the largest pool size were 13x13, 9x9, 5x5 and 1x1 (1x1 is no treatment), which can greatly increase the receptive field to separate the most significant context features. At the same time, yolov4 model can extract features from top to bottom repeatedly by using panet structure for three effective feature layers.

After the feature extraction of the target object, the yolov4 model uses the yolov3 head to predict the obtained features, and obtains the prediction results of three effective feature layers. The corresponding shape data is also obtained, and the position of the three prediction boxes can be determined. However, the prediction result does not correspond to the position of the final prediction frame on the original image, and each feature layer needs to be decoded. After decoding, the positions of prediction frames on the original image can be obtained, and these prediction boxes can be drawn on the original image after filtering by relevant algorithms.

The above process can complete the final rendering effect of yolov4 target detection.

2.2. Core Algorithm of Yolov4

The backbone network csparknet53 is the core algorithm of yolov4, which is used to extract target features. With reference to cspnet, CSP arknet53 can keep accuracy, reduce computing bottleneck and reduce memory cost while lightweight.

Yolov4 adds CSP to each large residual block of darknet53, divides the feature mapping of the basic layer into two parts, and then merges the cross stage hierarchies to ensure the accuracy while reducing the amount of calculation. The activation function of CSP arknet53 uses the mish activation function, while the back network uses leaky Relu function, the experimental results show that this setting in the target detection accuracy is higher.

Different from yolov3 algorithm using FPN for up sampling, yolov4 uses the idea of information flow in panet network for reference. Firstly, the semantic information of high-level features is transmitted to the low-level network through the above-mentioned sampling method, and then it is fused with the high-resolution information of the bottom features to improve the detection effect of small targets. Then the information transmission path from the bottom to the top is added, and the feature pyramid is strengthened by down sampling. Finally, the feature map fusion of different layers is used for prediction.

3. The Data Set of Power Inspection

3.1. Acquisition and Arrangement of Power Inspection Data Set

At present, the public data sets about power inspection are not disclosed because of confidentiality. After searching a lot of information, we only find the data sets about insulator images on GitHub. In order to obtain more abundant power inspection pictures in the scene, based on this experiment, we took the relevant scenes of power inspection and added some training pictures. Finally, 724 power inspection pictures are formed, some of which are shown in the figure below.
3.2. Data Annotation of Power Inspection

The data labeling tool used in this experiment is the open source project labelImg. The labeling method adopted is classified labeling. Common data labeling tasks include classification labeling, frame labeling, area labeling, trace point labeling and other labeling. Considering the business scenario of this experiment, the author uses the method of classification and labeling to label the above data.

Classification labeling is to select appropriate labels from a given label set and assign them to the labeled objects. The tags involved in this test are insulators, anti-vibration hammers, equalizing ring, link plates, bolts, spacers, houses, people, etc. After labeling with the open source tool labelImg, an xml markup file in PASCAL VOC format will be generated. The specific labeling example is as follows:

4. Target Detection Test of Power Inspection based on YOLOv4

4.1. Experimental Configuration

This experiment adopts the framework of Darknet deep learning. The test environment was Ubuntu 16.04, CUDA 10.1, cudn7.6.5 and python 3.7. The training settings are as follows:

- batch=16;
• subdivisions=8;
• max_ Batches is set to (classes * 2000), with a minimum of 4000. For example, if you train three target categories, max_ batches=6000;
• Changing steps to max_ 80% and 90% of batches, for example, steps = 4800, 5400;
• If you want to increase the network resolution, you can increase the value of height and width, but it must be a multiple of 32, which will help to improve the detection accuracy.

4.2. Experimental Analysis
The power inspection pictures in the above data set were marked by LabellImg visual picture annotation tool. When labeling, label different objects in the picture according to each label. The marked rectangular frame is mainly divided into eight categories: insulator, anti-vibration hammer, equalizing ring, connecting plate, bolt, spacer, house, and person. We selected 580 pictures from the above 724 pictures as the training set for this experiment, and the remaining 142 pictures were used as the test set for this experiment. After the training of the training set data is completed, the Precision and Recall of the target detection result of the power inspection should be tested. The specific expression is as follows:

\[
\text{Precision} = \frac{TP}{TP + FP} \quad (1)
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \quad (2)
\]

In the above formula, TP is used to indicate the number of correctly detected power inspection objects; FP is the number of falsely detected power inspection objects; FN is the number of missing power inspection objects.

The detailed test results are as follows:

|   | TP  | FP  | FN  |
|---|-----|-----|-----|
|   | 450 | 39  | 48  |

Table 1. Experimental effect parameter table.

Part of the detection effect on the test set is shown in the following figure. Different types of object images are marked with different color boxes and the corresponding object names are added. For example, for the insulator image, the model is marked with a red box and the word insulator is added; for the equalizing ring image, the model is marked with a light green box and the word Equalizing ring is added.

In the data set of this experiment, some pictures have overlapping phenomena of different objects, which directly reduces the detection results of the model. This model has a weak ability to recognize overlapping parts of objects. Finally, the detection accuracy rate is 0.92, and the recall rate is 0.904.
5. Summary
This experiment uses the YOLOv4 deep convolutional network as the technical means. After training the power inspection training data set sorted out above, the detection effect on the test data set is good, but there are mainly two problems:

There are fewer data annotation sets for power inspection. In this experiment, 580 images in the training data set were labeled. This number is far from meeting the requirements of YOLOv4 for data set training. This is an important source of the low accuracy of the power inspection target detection experiment.

The overlap of different objects in some photos of the data set is serious, which directly affects the recognition of different objects. If the overlapping area of different objects is too large, it will directly affect their recognition by the YOLOv4 algorithm. This explains the low recall rate of this trial.

6. References
[1] Bochkovskiy A, Wang C Y, Liao H Y M 2020 YOLOv4: Optimal Speed and Accuracy of Object Detection
[2] Wang C, Liao H M, Yeh I 2019 CSPNet: A New Backbone that can Enhance Learning Capability of CNN (arXiv: Computer Vision and Pattern Recognition)
[3] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S 2017 Feature pyramid networks for object detection (Proceedings of the IEEE conference on computer vision and pattern recognition) p 2117-2125
[4] Singh, B., & Davis, L. S 2018 An analysis of scale invariance in object detection snip (Proceedings of the IEEE conference on computer vision and pattern recognition) p 3578-3587
[5] Jun M R, Luo H B, Wang Z B 2019 Improved yolov3 algorithm and its application in small target detection (Acta Optica Sinica)
[6] Huang Z , Wang J 2019 DC-SPP-YOLO: Dense Connection and Spatial Pyramid Pooling Based YOLO for Object Detection (arXiv)
[7] Redmon, J., Farhadi 2018 Yolov3: An incremental improvement (arXiv preprint arXiv:1804.02767)
[8] Bozcan, I., Kayacan, E 2020 AU-AIR: A Multi-modal Unmanned Aerial Vehicle Dataset for Low Altitude Traffic Surveillance (arXiv preprint arXiv:2001.11737)
[9] Junming Chang, Qiong Ren, Hai Han and Lu Xu 2019 Integration and Service Strategy of VR/AR in Practical Teaching (IOP Conference Series: Materials Science and Engineering)
[10] Zou Z, Shi Z, Guo Y 2019 Object Detection in 20 Years: A Survey