Research on unmanned aerial vehicle power inspection technology based on YOLOV3

Jie Liu*, Weimo Lu, Zhikun Wang, Dong Xie, Hongbo Shi
Shaanxi Jingshen Railway Co., Ltd., Yulin Wealth Center, Yulin 719000, Shaanxi, China

ABSTRACT

In order to meet the requirements of real-time and accuracy for unmanned aerial vehicle (UAV) inspection of transmission lines, this paper deeply studies the application of YOLOV3 object detection algorithm in the onboard AI module of UAV inspection. By integrating the target detection candidate region selection and object recognition into one, the YOLOV3 algorithm, combined with multi-scale feature fusion, realizes high accuracy and real-time optimization of target detection and uses residual blocks to solve the problem of model degradation. The test results of transmission line insulators show that the average accuracy of YOLOV3 algorithm can reach 90%. Under the same conditions, the average processing speed of YOLOV3 algorithm is about 3.2 times that of Faster RCNN algorithm.

Keywords: Unmanned aerial vehicle, power line inspection, YOLOv3

1. INTRODUCTION

With the development of computer technology, target detection technology using UAV cruise combined with image recognition algorithm has been widely used in urban planning, emergency rescue, engineering construction, forest fire prevention, and power line inspection[1,2]. In power line inspection, drones equipped with AI modules can perform target detection and fault recognition on transmission lines, effectively reducing labor costs and labor intensity. The promotion and application of AI algorithms have significantly improved the automatic detection and judgment accuracy of transmission line defects and their real-time performance[3-5], further improving the efficiency of power transmission line inspection and ensuring the safe and stable operation of the power grid.

Scholars at home and abroad have extensively researched UAV image recognition and inspection technology[6,7]. Wang et al.[8] studied the application of convolutional neural network (CNN) in power inspection component detection by optimizing the extraction of target areas and improving the classifier method. Guo et al.[9] adopted a real-time target detection method based on deep learning algorithms, achieving target detection of various scales under various environments, with the highest detection speed reaching 30 frames/s. Pan et al.[10] analyzed the principle of scattering transform and convolutional neural networks, processed the scattering coefficients of low-pass filters, and used the Gram matrix algorithm to reduce noise interference, resulting in a 1.4% improvement in real-time positioning performance (recall rate) of insulator strings compared to the traditional SSD network framework.

The advantage of convolutional neural networks lies in the robustness of image recognition, but UAV inspection requires stable performance and real-time detection requirements[11,12]. Therefore, this paper studies the application of multi-rotor UAV AI modules with YOLOV3 image recognition network model in UAV power line inspection, optimizes the image recognition algorithm for accuracy and real-time requirements in power line inspection, further, improves the speed of image processing, and achievess a real-time return of detection results and automatic generation of detection reports.

2. AI MODULE ALGORITHM

2.1 YOLOV3 algorithm

Image processing and object detection technology are one of the key technologies for UAV inspection[13]. In order to achieve accurate positioning of inspection targets and rapid identification of defects, the YOLOV3 algorithm is selected for image recognition in the onboard AI module.

*154479652@qq.com

International Conference on Optics, Electronics, and Communication Engineering (OECE 2024), edited by Yang Yue, Proc. of SPIE Vol. 13395, 133951H · © 2024 SPIE · 0277-786X · Published under a Creative Commons Attribution CC-BY 3.0 License · doi: 10.1117/12.3050112

Drawing on the Faster RCNN prior box method, the target position coordinates of the YOLOV3 algorithm in the initial training stage are:

$$x = aw + x_0 \tag{1}$$

$$y = \beta h + \mathrm{y}_0 \tag{2}$$

wherein x and y are the center coordinates of the predicted bounding box of the algorithm; x_0 and y_0 are the center coordinates of the prior bounding box; w and h are the width and height of the prior bounding box; α and β are parameters to be learned. Since there is no restriction on the values of α and β, the center of the predicted bounding box may appear in any position when using Faster R-CNN for object detection, resulting in unstable algorithms during the initial training period. The YOLOV3 algorithm is adjusted based on this, and the center of the predicted bounding box is constrained within a specific grid, expressed as

$$x = \sigma(t_x) + d_x \tag{3}$$

$$x = \sigma\left(t_y\right) + d_y \tag{4}$$

$$w = w_0 e^{t_w} \tag{5}$$

$$h = h_0 e^{t_k} \tag{6}$$

In this equation, d_x and d_y represent the distance from the current grid's top left corner to the image's top left corner. w_0 and h_0 represent the initial width and height of the prior box. t_x, t_y, t_w, and th represent functions that need to be learned to predict the center coordinates, width, and height of the bounding box. The t_x and t_y are constrained within the range of (0, 1) using a σ function, making it easier for the model to learn and improving prediction stability.

The YOLOV3 network is based on the Darknet-53 feature extraction network for object classification, utilizing residual blocks to construct a multi-layer neural network classifier. The basic structure of the residual block is shown in Figure 1.

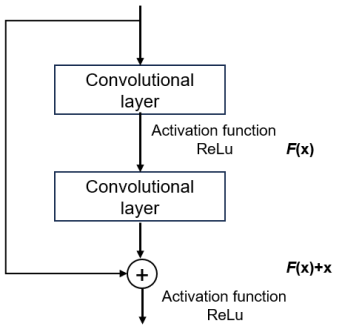

Figure 1. Schematic structure of residual block.

In Figure 1, x represents the input, and $F(x)$ represents the output after the first layer of linear transformation and activation. This process turns the feature functions, which were originally required to be learned independently layer by layer, into a process of filling and augmenting x, thereby reducing the learning difficulty.

2.2 Loss function

The loss function is the deviation between the actual output value of the network and the sample label. The YOLOV3 loss function generally consists of four parts, namely, the error of the center point of the bounding box, the error of the height and width of the bounding box, the error of the confidence, and the error of object classification. In essence, it uses the sum of the squares of the errors between the network output and each content of the sample label as the overall error of a sample.

The error of the center point of the frame can be expressed as

$$L_1 = \lambda \sum_{i=0}^{s^2} \sum_{j=0}^{B^2} I_{ij}[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \tag{7}$$

wherein: s^2 is the total number of grids; B is the total number of anchor boxes; λ is the error weight for bounding box regression, set to a value of 5; I_{ij} is used to indicate whether the detection object is included in the error. If the jth anchor

box in grid i contains the target object, then I_{ij} is set to 1, and the bounding box regression error is included in the comprehensive error. If the jth anchor box does not contain the target object, then I_{ij} is set to zero, and no error data for this item is recorded.

Similarly, the error of the frame width and height can be expressed as

$$L_2 = \lambda \sum_{i=0}^{s^2} \sum_{j=0}^{B} I_{ij} [(\sqrt{w_i} - \sqrt{\widehat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\widehat{h}_i})^2] \tag{8}$$

By taking the square root of height and width, the difference in sensitivity among different objects is reduced in error calculation, making them have similar weights in dimensional errors.

The confidence error function can be composed of two parts.

$$L_3 = \lambda \sum_{i=0}^{x^2} \sum_{j=0}^{B} I_{ij} (C_i - \hat{C}_i)^2 + \lambda' \sum_{i=0}^{s^2} \sum_{j=0}^{B} I'_{ij} (C_i - \hat{C}_i)^2 \tag{9}$$

wherein: λ' is the confidence weight of the frame when there is no detection object, and this paper takes 0.5 as a rule of thumb; I'_{ij} represents the parameter of no detection object, and its value is opposite to I_{ij}; Ci is the confidence of the presence of a detection object in the ith grid. The first term in equation (9) represents the confidence when there is a detection object in the cell, and the prediction of the regression confidence of the frame needs to include the error. The second term is used to describe confidence when there is no detection object in the cell.

Object classification error can be expressed as

$$L_4 = \sum_{i=0}^{s^2} I_i \sum (p_i(c) - \hat{p}_i(c))^2 \tag{10}$$

where: I_i represents that only the grid where the detection object exists will record this type of error; $P_i(C)$ is the classification error value of the detection object in the ith grid.

In summary, the prediction integrated error can be characterized by the sum of the above four error functions, that is

$$L = L_1 + L_2 + L_3 + L_4 \tag{11}$$

2.3 Optimization of application

For the power devices and defect datasets, the K-means++ algorithm is used to ensure the discrete distribution of initial points to enhance the rationality of the prior frame. When calculating, first randomly determine a cluster center from the training set target frames, then calculate the Euclidean distance $D(s)$ between each sample s and the cluster center, and calculate the probability through equation (12), selecting the sample point corresponding to the maximum value as the next cluster center.

$$\eta = \frac{D^2(s)}{\sum_{s \in S} D^2(s)} \tag{12}$$

The process is repeated according to the roulette method. Considering the changes in image size caused by the changes in the angle of the detection target, in order to cover different sizes of detection targets, this paper selects nine prior frames with sizes of 10×13, 16×30, 33×23, 30×61, 62×45, 59×119, 116×90, 156×198, and 373×326 as the cluster centers. After obtaining nine cluster centers from the training set, the samples in the target frames need to be attributed to the clusters corresponding to the cluster centers. The cluster centers are recalculated according to the above process until the cluster centers no longer change, and then the prior frame cluster centers that meet the requirements are obtained.

3. MODEL EVALUATION AND ANALYSIS

In object detection, it is necessary to evaluate the location and classification of the detected objects. The evaluation metrics include precision and recall, and the calculation formulas are as follows

$$P_r = \frac{T_P}{F_P + T_P} \tag{13}$$

$$R = \frac{T_P}{F_P + F_N} \tag{14}$$

where T_P is true positive, F_P is false positive, F_N is false negative. In order to describe the model performance more accurately, the average precision is used to evaluate the model accuracy. For each different recall rate value (0 to 1, every 0.1, a total of 11 points), the maximum precision rate when it is greater than or equal to these recall rate values is selected, and then the average of the precision rates of all categories is calculated as the average precision, which is expressed as

$$P_a = \frac{1}{11}\sum_{R\in\{0,0.1,\ldots,1\}} P'(R) \tag{15}$$

$$P'(R) = \max P(\hat{R}) \tag{16}$$

wherein: $P'(R)$ is the maximum precision when the recall rate satisfies $R \geq R$; R is the recall rate when the maximum precision is achieved.

The application effects of three target detection algorithms, namely Faster RCNN, SSD, and YOLOV3, in recognition of transmission line insulators, were analyzed and compared. A total of 600 inspection images were used, including 300 images of insulators with defects and 300 images of insulators without defects. The GPU model used was an NVIDIA GTX 1080Ti/1GB. Figure 2 shows the comparison of detection results among different algorithms. From Figure 2, it can be seen that the average accuracy of both Faster RCNN and YOLOV3 algorithms is above 90%, while the average accuracy of the SSD algorithm is about 80%.

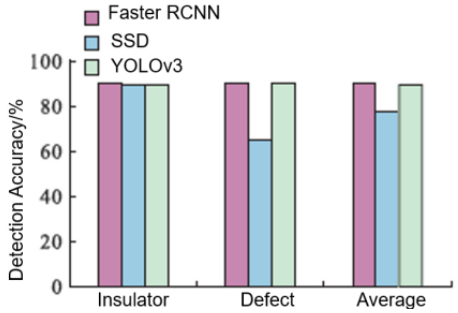

Figure 2. Comparison of detection results for three algorithms.

To further verify the applicability of the YOLOV3 algorithm, 11,942 power inspection images were selected as a training set, and 1,916 images were used as a validation set for object detection and fault recognition. The detection results are shown in Table 1.

Table 1. Detection results for different types of defects.

Type	P_a	R	P_r
Glass cesium edge carrier	89.52	92.79	85.56
Composite insulators	86.63	89.58	88.73
Ceramic insulators	88.49	90.38	90.18
Connection fittings	81.14	86.54	92.59
Hanging point fittings	82.17	88.59	93.50
Shockproof hammer	95.35	96.90	89.17
Equalizing ring	80.94	88.16	94.61
Insulator self-explosion	79.92	82.69	86.72
Damaged anti-hair hammer	80.65	83.08	83.17

Table 1 shows that the YOLOV3 algorithm is widely applicable to different detection targets and common fault types. The average precision of each detection target is higher than 79%, and the recall rate is greater than 82%.

4. CONCLUSION

This article investigates the application effect of the YOLOV3 object detection algorithm in the airborne AI module of UAV inspection and reaches the following main conclusions:

(1) Taking the detection of transmission line insulators as an example, the average accuracy of the YOLOV3 algorithm exceeds 90%; under the same conditions, the average processing speed of the YOLOV3 algorithm is approximately 3.2 times that of the Faster RCNN algorithm and approximately 1.6 times that of the SSD algorithm.

(2) For different detection objectives and common fault types, the average accuracy of the YOLOV3 algorithm is higher than 79%, and the recall rate is greater than 82%.

REFERENCES

[1] Huang, X., et al., "Classification recognition method of insulator in aerial image based on the red-blue difference and developed K-means algorithm," High Volt. Eng. 44, 1528-1534 (2018).

[2] Feng, Z. H., Zhang, X. F., Fang, S. B., et al., "Research on route planning of UAV inspection based or3D GIS," High Voltage Apparatus 53(8), 81-86 (2017).

[3] Yang, W., Zhou, H., Yang, S. L., et al., "Transmission line inspection insulator tracking and ranging technology," Electronic Science and Technology 30(5), 124-127 (2017).

[4] Wang, S. B., Li, M. X., et al., "Research on obstacle detection of transmission line corridor based on 3D laser radar technology," Electronic Science and Technology 32(4), 81-84 (2019).

[5] Ren, P. F., "WSN node localization algorithm for power transmission networks based on particles warm optimization," Journal of Shenyang University of Technology 40(5), 541-546 (2018).

[6] Wang, S. B., et al., "The design of unmanned aerial vehicle power patrol base on oblique photography technology," Electronic Science and Technology 32(5), 85-88 (2019).

[7] Feng, M., Luo, W., Yu, L., et al., "A bolt detection method for pictures captured from an unmanned aerial vehicle in power transmission line inspection," Journal of Electric Power Science and Technology 33(4), 135-140 (2018).

[8] Wang, W. G., Tian, B., Lu, Y., et al., "Study on the electrical devices detection in UAV images based on region based convolutional neural networks," Geoinformation Science 19(2), 256-263 (2017).

[9] Guo, J. D., Chen, B., Wang, R. S., et al., "YoLo-based real-time detection of power line poles from unmanned aerial vehicle inspection vision," Electric Power 52(7), 17-23 (2019).

[10] Pan, C., Shen, P. F., Zhang, Z., et al., "Research on real-time positioning of insulator strings based on UAV inspection images," Insulators and Surge Arresters 1, 234-240 (2020).

[11] Chang, W. Z., et al., "Research of unmanned aerial vehicle comprehensive inspection for distribution network overhead transmission lines," Electric Power 51(1), 97-101 (2018).

[12] Mao, T. Q., Wang, C. C., et al., "Transmission line extraction method based on SGC stereo matching algorithm for UAV inspection image under complex background," Smart Power 46(12), 105111 (2018).

[13] Feng, Z. H., Wang, L. X., Liang, W. Y., et al., "Research on image information acquisition method of typical transmission line equipment based on UAV," Insulators and Surge Arresters 4, 2226 (2016).