Bird Detection on Transmission Lines Based on DC-YOLO Model

Cong Zou, Yong-Quan Liang

Bird Detection on Transmission Lines Based on DC-YOLO Model

Cong Zou[1] ,Yong-quan Liang[1,2,*]

[1] College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, China
[2] Provincial Key Laboratory for Information Technology of Wisdom Mining of Shandong Province, Shandong University of Science and Technology, Qingdao, China
zc17860763660@163.com

Abstract. In order to accurately detect the number of birds around the transmission line, promptly drive the birds away to ensure the normal operation of the line, a DC-YOLO model is designed. This model is based on the deep learning target detection algorithm YOLO V3 and proposes two improvements: Replacing the convolutional layer in the original network with dilated convolution to maintain a larger receptive field and higher resolution, improving the model's accuracy for small targets; The confidence score of the detection frame is updated by calculating the scale factor, and the detection frame with a score lower than the threshold is finally removed. The NMS algorithm is optimized to improve the model's ability to detect occluded birds. Experimental results show that the DC-YOLO model detection accuracy can reach 86.31%, which can effectively detect birds around transmission lines.

Keywords: Bird Detection, Deep Learning, YOLO V3, Dilated Convolution, NMS Algorithm.

1 Introduction

Transmission lines play a pivotal role in the power system, and their construction scale has increased dramatically in recent years. Bird damage is an important factor that threatens the safety of the line. Its impact on the transmission line mainly has four aspects: bird pecking, bird nesting, bird excretion, and bird flight. Aiming at the problem of bird damage, the current effective method is to install an ultrasonic bird repellent, but long-term work of the bird repellent will cause waste of energy consumption. Therefore, it is extremely important to implement accurate detection of transmission lines and timely start bird repellents when a certain number of birds are moving around the lines.

In recent years, deep learning [1-3] technology has been widely used in practical scenarios. Deep convolutional neural networks can learn autonomously when performing target detection. Target detection algorithms based on deep learning can be roughly divided into two categories: 1) two stage target detection algorithms, such as Fast R-CNN [4], Faster R-CNN [5], etc. This kind of algorithm is tested in two steps , first use

Region Proposal Network (RPN) [6] to generate candidate areas, and then achieve target detection classification; 2) one stage target detection algorithms, such as SSD [7], YOLO V3 [8,9], etc. Such algorithms directly predict the position of the target via the detection network and category information, which has faster detection speed and can basically achieve real-time detection.

Based on the detection algorithm YOLO V3, this paper proposes DC-YOLO model, and improves the structure of YOLO V3 in response to the problems of small targets and mutual occlusion of birds in the device acquisition pictures. In order to improve the recall rate and precision rate of small targets in the image by the network, the convolutional layer is replaced with dilated convolutional layer [10-13] to maintain a large receptive field and a higher resolution. According to the Intersection-over-Union (IOU) [14] value of the detection frame and the pre-selected detection frame, calculate the scale factor corresponding to each detection frame, thereby attenuating their confidence scores, and finally iteratively delete the detection frames whose score is lower than the set threshold. The improved network in this paper is compared with various networks on the transmission line bird data set. The experimental results show that the improved network has better detection effect.

2 YOLO V3 Algorithm

Based on YOLO V2, YOLO V3 combines the ideas of ResNet [15], FPN [16], and binary cross-entropy loss. Its backbone network is composed of 53 consecutive 1×1、3×3 convolution layers, called Darknet-53 [17]. The structure is shown in Figure 1. The network outputs features at three scales of 13×13、26×26, and 52×52. The detection network performs regression analysis on the features of the three scales to generate multiple prediction frames. The NMS algorithm [18-20] removes redundant prediction frames and retains the prediction frame with a higher confidence score as the target detection frame.

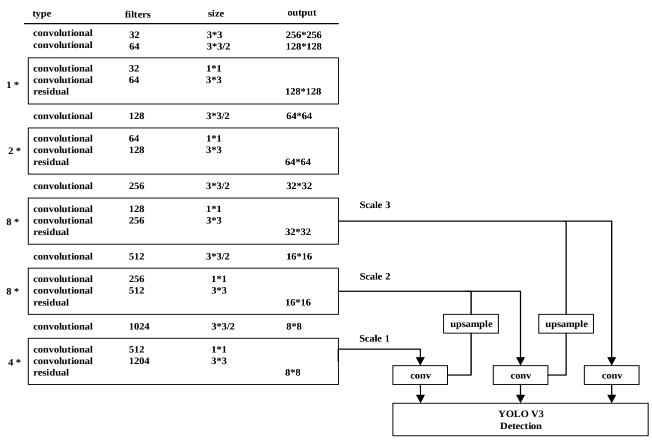

Fig. 1. YOLO V3 network structure

The YOLO V3 network first scales the input image to 416×416 and divides the image into 13×13 cells. If there is a target object in the center of a cell, the cell is responsible for detecting this object. Each cell will generate A prediction frames, the prediction frame consists of a five-dimensional prediction parameter, including the coordinates of the center point (x,y), the width and height (w,h), the confidence score s_i, the confidence score is calculated by Equation (1).

$$s_i=P(C_i|O_{object})\times P(O_{object})\times IOU(truth,pred) \tag{1}$$

where $P(O_{object})$ represents the possibility of an object in the current cell detection frame, if there is an object to be detected, the value is 1, otherwise, the value is 0. $P(C_i|O_{object})$ means the conditional probability that the cell predicts the i type object when there is an object in the detection frame. $IOU(truth,pred)$ is the intersection ratio of the predicted detection frame and the real labeled frame.

Finally, the prediction frame with a higher confidence score is retained as the target detection frame by the NMS algorithm. The traditional NMS processing method is expressed in Equation (2):

$$s_i=\begin{cases} s_i, & IOU(M,b_i)<N_t \\ 0, & IOU(M,b_i)\geq N_t \end{cases} \tag{2}$$

where M is the prediction frame with the largest confidence score in the current region. $IOU(M,b_i)$ is the intersection ratio of M and adjacent overlapping frame b_i. N_t is the set overlap threshold.

3 DC-YOLO Model

3.1 DC-YOLO Backbone Network Structure

In order to make the network have a wider receptive field, a dilated convolution is added to the DC-YOLO structure. The dilated convolution increases the receptive field of the convolution kernel by changing the internal rate of the convolution kernel. Figure 2 is a comparison diagram of dilated convolution kernels at different rate. (a) the figure corresponds to a 3×3 dilated convolution with rate=1, that is a standard convolution, at this time, the convolution kernel receptive field range is 3×3. (b) the graph corresponds to a 3×3 dilated convolution with a rate=2, the actual convolution kernel size is still 3×3, but for a 7×7 image only 9 points have a convolution operation, which can be understood that the weights of 9 points in the picture are not 0, and the rest are 0. So when rate=2, the receptive field of the convolution kernel increases to 7×7. Therefore, the use of dilated convolution can increase the receptive field without increasing the amount of parameters, so that the convolutional network can extract feature information of a larger field of view.

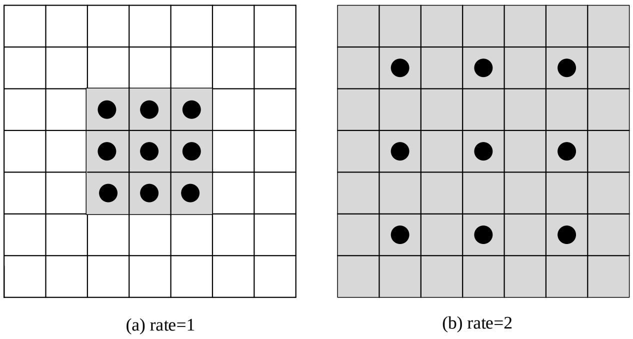

Fig. 2. Dilated convolutions.

In the YOLO V3 structure, the small target semantic features in the 13×13 scale feature map are seriously lost, and the 52×52 scale feature map has higher resolution, but it will cause larger calculation and memory overhead. In DC-YOLO, the network's last two downsampling of the input image is cancelled, so that the resolution of 26×26 is maintained in the last three stages. This can not only ensure moderate computing overhead, but also increase the resolution of the final output feature map, reduce the loss of semantic features of small-sized targets in deep networks, and facilitate the detection of small targets. Because reducing the number of times of image downsampling will inevitably reduce the receptive field of the deep network, the DC-YOLO structure will introduce two types of dilated convolution residual structure A and B as shown in Figure 3.

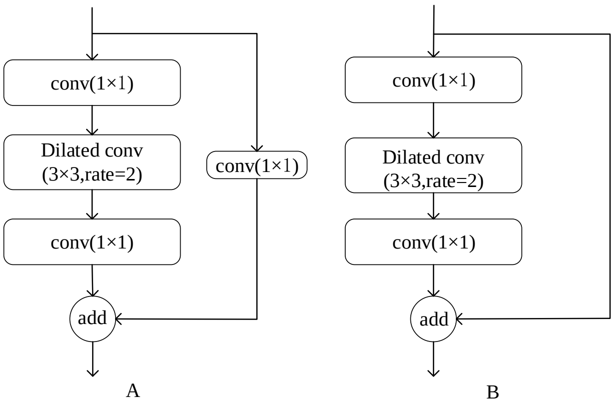

Fig. 3. Structure of dilated convolution residual

The DC-YOLO backbone network is shown in the figure 4. The resolution of the features directly affects the detection of small targets. The model selection has a moderate 26×26 resolution, which enables the model to maintain higher resolution and larger receptive field in the deep network structure, to enhance the ability to detect small targets.

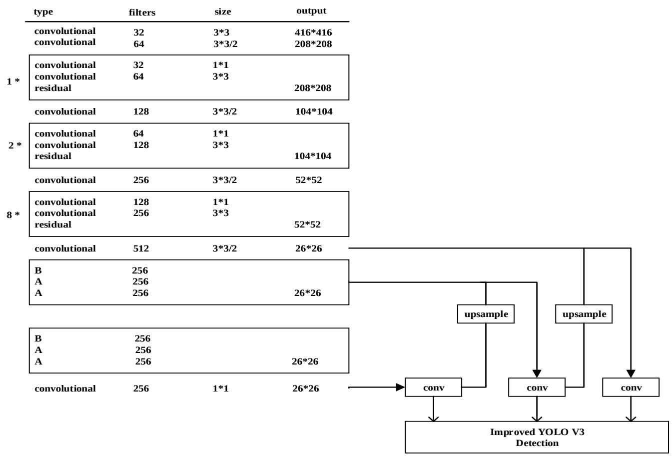

Fig. 4. Improved YOLO V3 network structure

3.2 Scale factor NMS Algorithm

When the traditional NMS algorithm suppresses redundant prediction frames, the judgment of whether a prediction frame is redundant mainly depends on the size of the overlap threshold. The algorithm forcibly sets the confidence score of prediction frames larger than the overlap threshold to 0. When the real target appears in the overlapping area, it will be deleted by mistake, which easily leads to missed detection of the target.

In the bird detection task of transmission lines, bird gathering often occurs. In this paper, the NMS algorithm based on scale factor attenuation is applied to the detection task, and the confidence score of the prediction frame is attenuated according to a certain proportion according to the degree of overlap, so that the algorithm can effectively suppress the redundant prediction frame and reduce the missed detection rate of the target. The algorithm is as follows: b_m is the prediction frame with the highest confidence score in the prediction frame set B. Calculate the $IOU(b_m,b_i)$ value of the remaining adjacent prediction frames and b_m. Based on this value, use Equation (3) to obtain the scaling factor w_i of each prediction frame, and the confidence score of the attenuated frames is $w_i s_i$. Finally, delete the prediction frames whose confidence scores are less than the set threshold after attenuation, and repeat this process until all the prediction frames in set B have been processed. This algorithm calculates the scale factor corresponding to the prediction frame according to the value of *IOU*. It is a continuous process. When the value of *IOU* is 0, the original confidence score of the prediction frame is retained.

$$w_i = 1 - lg(IOU(b_m,b_i)+1) \tag{3}$$

4 Experimental Results and Analysis

4.1 Experimental Data Set and Preprocessing

The data used in the experiment mainly comes from the monitoring equipment near the transmission line in a certain area collected by the research team. Randomly extract the monitoring video and extract a single frame image to make a data set. Label the picture with Labbellmg tool and store it in VOC data format. There are a total of 6000 images, including 4800 in the training set and 1200 in the test set.

According to the characteristics of the research object in this article, combined with the relationship between the number of prior frames and the average Intersection-over-Union (Avg IOU), as shown in the Figure 5. K-means clustering analysis is used to obtain the best prior frame. The distance measurement formula is:

$$d(box,centroid)=1-IOU(box,centroid) \tag{4}$$

where *box* is the sample; *centroid* is the cluster center; *IOU*(*box*,*centroid*) is the intersection ratio of the cluster center and the sample frame.

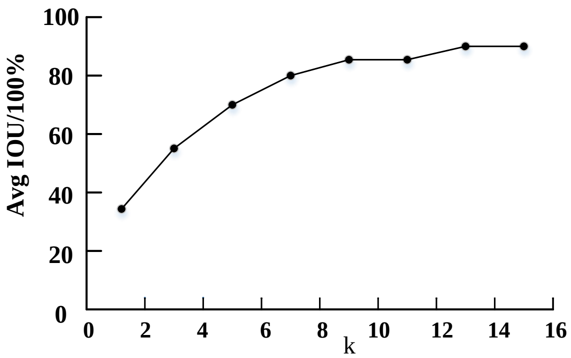

Fig. 5. K-means clustering result

When the value of k is 9, the curve gradually starts to flatten, so the number of anchor boxes is 9 and the size of the corresponding prediction frame is set to 9 cluster centers. On this training set, they are: (10,13), (16,28), (30,33), (33,61), (64,50), (69,110), (115,90), (155,198), (369,342).

4.2 Network Training

Model training parameters are shown in Table 1. In the initial stage of training, the learning rate is 0.001 to stabilize the network. When the number of training iterations is 20000, the learning rate is adjusted to 0.0001, and when the number of iterations is 25000, the learning rate is adjusted to 0.00001, which further converges the loss function.

Table 1. Model parameter description

Parameter	Value
Batch Size	64
Max batches	30000
Momentum	0.9
Decay	0.0005
Match Threshold	0.25
NMS threshold	0.5

4.3 Results Analysis

In order to verify that the improved network can better detect small targets and birds obstructing each other, thereby improving the overall detection effect, two comparative experiments are set up in this paper.

Small Target Detection.
The SSD, Faster-RCNN, YOLO V3, and the DC-YOLO model were trained and tested on the dataset respectively. Calculate the *Precision* and *Recall* of the target, the *P* refers to the proportion of the number of correctly detected targets in the detection result, *R* refers to the proportion of the number of correct detection results to the total number of targets to be detected. They are defined as follows:

$$P=\frac{X_{TP}}{X_{TP}+X_{FP}} \tag{5}$$

$$R=\frac{X_{TP}}{X_{TP}+X_{FN}} \tag{6}$$

where X_{TP} is the number of targets detected correctly; X_{FP} is the number of targets detected incorrectly; X_{FN} is the number of targets not detected.

The experimental results are shown in Table 2. The DC-YOLO model improves the accuracy and recall of detection targets to varying degrees.

Table 2. Detection results of different models

Detection model	*P*(%)	*R*(%)	*AP*(%)
SSD	73.35	74.52	73.93
Faster-RCNN	79.68	82.17	80.91
YOLO V3	76.23	77.85	77.03
DC-YOLO	84.57	88.12	86.31

The average precision (AP) measures the accuracy of the model from two angles of *Precision* and *Recall*. It can be used to analyze the detection effect of a single category. The AP curve of the DC-YOLO model is shown in the Figure 6, the AP value is increased to 86.31%.

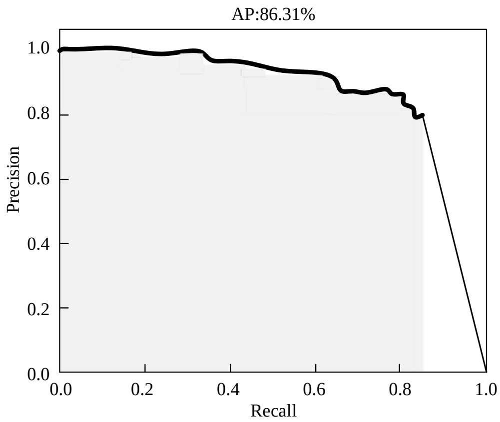

Fig. 6. AP curve of DC-YOLO

Detection of Mutual Occlusion of Birds.

In the experimental data set, 100 bird images with different occlusion ratios were selected, and they were detected using the traditional NMS algorithm and the scale factor NMS algorithm improved in this paper. The detection performance is shown in Table 3. When the bird's mutual occlusion ratio is less than 30%, the detection accuracy of the two algorithms is the same; when the occlusion ratio is 30%-70%, the scale factor NMS algorithm improves the accuracy rate by 20%; When the occlusion ratio is higher, the scale factor NMS algorithm shows better performance in bird occlusion detection tasks.

Table 3. Comparison of detection performance with different occlusion ratios

Detection algorithm	Occlusion ratio(%)	AP(%)
Traditional NMS algorithm	0-30	90
	30-70	40
	≥70	10
Improved NMS algorithm	0-30	90
	30-70	60
	≥70	20

Figure 7 shows the detection results of the traditional NMS algorithm and the improved NMS algorithm at a occlusion ratio of 80%. It can be seen that the traditional algorithm cannot detect the blocked birds, and the improved NMS algorithm has a better detection effect on the blocked birds .

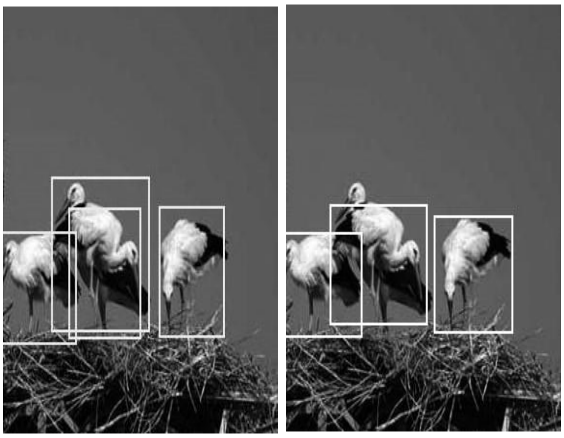

Fig. 7. Comparison of detection results between improved NMS algorithm and traditional NMS algorithm

5 Conclusion

This paper proposes a DC-YOLO model to complete real-time and effective detection of birds on transmission lines, thereby controlling the start and stop of bird repellents, and protecting the stability of transmission lines while saving energy. Aiming at the small target of the image and the problem of target occlusion, based on the YOLO V3 model, the backbone network Darknet-53 network and the NMS algorithm were improved. The experimental results show that the DC-YOLO model has a higher precision rate, and the average detection speed reaches 38 FPS, which achieves a real-time accurate detection effect.

References

1. Le Cun Y, Bengio Y, Hinton G. Deep learning. Nature, 2015, 521(7553):436-444.
2. GOODFELLOW I, BENGIO Y, COURVILLE A. Deep Learning. Cambridge, UK:MIT Press, 2016:1-33.
3. NORDENG I E, HASAN A, OLSEN D. DEBC detection with deep learning. Scandinavian Conference on Image Analysis, Troms, Norway: Springer, 2017:248-259.
4. Girshick R. Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision. 2015:1440–1448.
5. Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.
6. Girshick R, Donahue J, Darrell T, et al.Rich feature hier-archies for accurate object detection and semanticsegmentation. IEEE Conference on Computer Visionand Pattern Recognition, 2014：580-587.
7. Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot Multi Box Detector. European Conference on Computer Vision.Springer International Publishing, 2016: 21-37.

8. Tumas P, Serackis A. Automated Image Annotation based on YOLOv3.2018 IEEE 6th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE). IEEE, 2018: 1-3.
9. Qu H, Yuan T, Sheng Z, et al. A Pedestrian Detection Method Based on YOLOv3 Model and Image Enhanced by Retinex.2018 11th International Congress on Image and Signal Processing, Bio Medical Engineering and Informatics (CISP-BMEI). IEEE, 2018:1-5.
10. Yu F, Koltun V, Funkhouser T. Dilated residual networks.Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 472-480.
11. Yanjie Wang, Guodong Wang, Chenglizhao Chen, Zhenkuan Pan. Multi-scale dilated convolution of convolutional neural network for image denoising. Multimedia Tools and Applications, 2019, Vol.78 (14), pp.19945-19960.
12. Duc My Vo, Sang-Woong Lee. Semantic image segmentation using fully convolutional neural networks with multi-scale images and multi-scale dilated convolutions. Multimedia Tools and Applications, 2018, Vol.77 (14), pp.18689-18707.
13. Geng Lei, Zhang Siqi, Tong Jun,Xiao Zhitao. Lung segmentation method with dilated convolution based on VGG-16 network. Computer assisted surgery (Abingdon, England),2019,24(sup2).
14. Bodla N, Singh B, Chellappa R, et al. Soft-NMS--Improving Object Detection With One Line of Code. Proceedings of the IEEE International Conference on Computer Vi-sion. 2017: 5561-5569.
15. He K, Zhang, X, Ren S, et al. Deep Residual Learning for Image Recognition.IEEE Conference on Computer Vision and Pattern Recognition ,Las Vegas, 2016:770-778.
16. Lin T Y, Dollár P, Girshick R, et al.Feature pyramid networks for object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Rec-ognition，2017：2117-2125.
17. Redmon J, Farhadi A.YOLOv3：an incremental improve-ment.IEEE Conference on Computer Vision and Pattern Recognition, 2018:89-95.
18. Rosenfeld A, Thurston M, Edge and curve detection for visual scene analysis. IEEE Transactions on Computers, 1971, 20(5):562-569.
19. Hosang J, Benenson R, Schiele B. A convent for non-maximum suppression. Proceedings of 38th German Conference on Pattern Recognition. Hannover, Germany:Springer, 2016:192-204.
20. Hosang J, Benenson R, Schiele B. Learning non-maximum suppression. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA:IEEE, 2017:6469-6477.