Foreign object detection of transmission line based on improved Yolov5

Gengwu Wei*[,a], Yingna Li[b]
[a] Faculty of Information Engineering and Automation, Kunming University of Science and Technology, 727 Jingming South Rd.,Chenggong District, Kunming, China 650500; [b] Computer Technology Application Key Lab of the Yunnan Province, 727 Jingming South Rd.,Chenggong District,Kunming, China 650500.

ABSTRACT

The environment of the transmission line is complex and easy to attach foreign matter, which has always been one of the reasons for the safety hazards to the operation of the transmission line. In view of the current problem that the detection accuracy of the foreign matter on the transmission line needs to be improved, an improved transmission line foreign matter detection method based on the Yolov5 algorithm is proposed. Detection model. First, the RepVGG module is introduced into the feature extraction network to enhance the network feature extraction ability and improve the model reasoning speed; secondly, the ability of the network to identify important feature information is strengthened by integrating the attention mechanism module; finally, by adding the prediction layer and Soft-nms The algorithm processes the target prediction frame to improve the detection accuracy of the model. The experimental results show that the improved Yolov5 transmission line foreign object detection algorithm proposed in this paper has a mAP value of 4.1% higher than the conventional Yolov5, and it also has certain advantages in performance compared with the conventional target detection algorithm.

Keywords: Foreign object detection, Yolov5, RepVGG, CBAM

1. INTRODUCTION

The demand for energy is increasing as human society starts to develop at a rapid pace. Since the second industrial revolution, electricity has always been an important force in the development of human society. Various foreign objects are also prone to appear on transmission lines, such as kites, balloons, and bird's nests. Therefore, it has become an urgent problem for the power sector to take relevant and effective measures to solve the problem of foreign objects on wires.

The foreign object inspection methods of transmission lines are primarily divided into manual inspection, deep learning-based detection techniques and traditional image detection techniques. Traditional image detection methods can achieve better results in scenarios with obvious features and uncomplicated backgrounds, but in reality, the targets to be detected are often complex and changeable, and the detected areas are not targeted, resulting in high time complexity of the algorithm and low detection efficiency. Due to the small number of professional inspection personnel in manual inspection, the inspection process is time-consuming and laborious, and there are great safety hazards in the detection of climbing poles. Deep learning is being developed, a large quantity of data is used to train the model, and feature extraction is performed in various complex situations, so that the algorithm has better generalization ability in various application scenarios, and the application in image processing has gradually become mainstream. The deep learning detection algorithms are broadly classified into two types. One is to generate a region proposal (Region Proposal) based on the picture, and then send the region proposal to the classifier for classification, which is completed by two different network structures. This target detection algorithm is called Two-stage detection model, represented by RCNN[1] 、 Faster-RCNN[2]. The other is to directly use a network structure to predict the classification and location coordinates of the intended object based on the input picture. A single-stage model is the name given to this type of object recognition method, represented by the Yolo[3-6] series and SSD[7]. In comparison to the two-stage model, a suggestion area is not required for the single-stage target detection model, and the detection speed is fast, so it is more suitable for rapid detection of the recognition target in a real-time environment.

*6837401@qq.com; [1]2032727859@qq.com

Seventh International Conference on Mechatronics and Intelligent Robotics (ICMIR 2023), edited by Srikanta Patnaik, Tao Shen, Proc. of SPIE Vol. 12779, 1277928 · © 2023 SPIE · 0277-786X · Published under a Creative Commons Attribution CC-BY 3.0 License · doi: 10.1117/12.2689392

2. RELATED WORK

Yolov5 is selected as the baseline algorithm, and an improved Yolov5 transmission line foreign object detection model is proposed. The improved network framework structure is seen in Figure 1. First, the RepVGG[8] module is integrated into the Yolov5 feature extraction network to reduce the loss of feature information and quicken up model reasoning speed. secondly, the mixed attention mechanism module CBAM[9] is made available to improve the network's ability to express features; The detection layer is then added to improve the detection ability of targets of various scales.Finally the Soft-NMS[10] algorithm is used to replace the NMS algorithm in order to reduce the likelihood of the model missing detection .

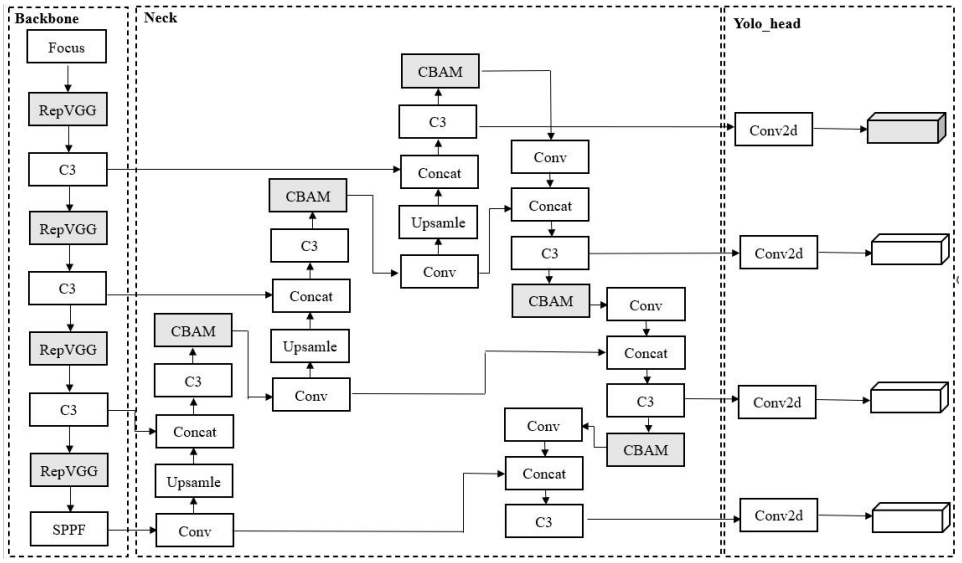

Figure 1. Improve the network structure of Yolov5

2.1 Introduction to YOLOv5 algorithm

The Yolov5 model is separated into five versions based on the depth and width of the network, namely Yolov5x, Yolov5l, Yolov5m, Yolov5s, Yolov5n. Due to the high real-time performance requirements of the model for foreign object recognition on transmission line towers, this paper considers the model's identification accuracy and inference speed, and chooses Yolov5s with a small model weight as the basic model for research.

The four key sections of Yolov5s are input, backbone, neck, and prediction. The input side adopts the Mosaic algorithm, and randomly selects four pictures in the data set to stitch together after cropping, zooming, flipping and other operations to increase the robustness of the model. Backbone integrates Focus, CSPNet, and SPPF modules. Focus slices the input image to lessen the amount of information lost during the downsampling process. The CSPNet structure divides the original input into two parts, one branch performs the Bottleneck×N operation, the other branch directly performs the convolution operation, and then performs the concat operation on the two branches to reduce quantity of channels, so that the BottleneckCSP's output and input. The same size so that the model can learn features better. The SPPF module uses multiple small-size pooling kernel cascades to replace the SPP module contains a large-size pooling kernel to fuse feature maps of different receptive fields, improve feature expression capabilities and speed up operation. The Neck part is mainly composed of FPN and then PANet network. FPN constructs high-level semantic feature maps at all scales via top-down side connections. A bottom-up path is introduced in the PANet network to make the underlying features better. The transfer to the top layer to combine the features of different layers improves the utilization of low-level features. The bounding box loss function used by the prediction layer is CIOU loss, and redundant prediction boxes are filtered using the non-maximum suppression approach.

2.2 RepVGG module

Although complex multi-branch models such as ResNet can achieve high accuracy, they also have certain defects. This will lead to a decrease in the calculation speed of the model and a decrease in memory usage. Some nodes will increase memory consumption, have high requirements for equipment, and increase hardware overhead. RepVGG is an improved

classification network based on the VGG network, as shown in Figure 2. Its main features are that the Identity and residual branches are added to the Block module of the VGG network, and in the model reasoning stage, 3×3 、 1 ×1 convolution branch, which facilitates the deployment of the model and speeds up the reasoning.

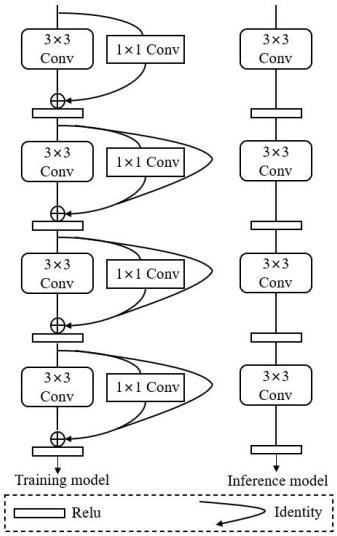

Figure 2. RepVGG training model and inference model structure

2.3 Hybrid attention system

Introducing the convolutional neural network's attention mechanism can reduce the negative impact of complex information when recognizing objects, that is, focus on key information and suppress irrelevant data to increase the model's resilience and ability to recognize objects. CBAM is a lightweight attention mechanism module that combines space and channels, and can be integrated into the convolutional neural network architecture with only a small computational overhead. Input the feature map's dimensions $F \in R^{C \times H \times W}$ to the CBAM module, After the input feature map undergoes two different pooling average pooling and maximum pooling, the two one-dimensional tensors obtained by pooling are obtained by sharing the layer that is totally connected and the corresponding activation function to obtain the correlation between channels, and the two output Merge to get the weight of each channel with features to generate channel attention features $M_c \in R^{1 \times 1 \times C}$, The mathematical expression of channel attention Mc is seen in formula (1).

$$M_c(F) = \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{\max}^c))) \tag{1}$$

The spatial attention module receives its input from the channel attention module's output, or Mc, and the procedures of global maximum pooling and global average pooling are respectively performed along the channel dimension to obtain a two-dimensional feature map $F_{avg} \in R^{1 \times H \times W}$, $F_{max} \in R^{1 \times H \times W}$ and perform dimension splicing on the channel to obtain an efficient feature description, perform convolution operation on the concatenated results, and output the map of spatial attention Ms after the Sigmoid activation function, The mathematical expression of channel attention Mc is seen in formula (2).

$$M_s(F) = \sigma(f([AvgPool(F); MaxPool(F)])) \tag{2}$$

2.4 Improved non-maximum suppression

A crucial step in the target detection process is non-maximum suppression (NMS). The primary concept is to rank the candidate boxes by confidence scores from high to low, and choose the candidate box with the most assurance, and the remaining candidate boxes. Perform IoU comparison and filter out boxes that exceed a set threshold of IoU. The NMS calculation formula(3) is as follows, s_i represents the confidence score of the i candidate frame, M is the candidate frame with the highest confidence score, N_t is the set IoU threshold, and b_i is the rest of the candidate frames. The most issue with NMS is to forcibly remove all candidate boxes whose IoU exceeds the threshold. The detection of an object will be unsuccessful if it occurs in the overlapping area, resulting in missed detection, and reduce the recall rate of the model.

$$s_i = \begin{cases} si, iou(M, bi) \geq N_t \\ 0, iou(M, bi) < N_t \end{cases} \tag{3}$$

In the scene of a real transmission line, foreign objects on the transmission line are often irregular in shape and easily form mutual occlusion with the tower. For this problem, it is a more effective method to use the confidence decay function to adjust the overlapping frame with the candidate frame with the highest confidence. The Soft-NMS algorithm is different from NMS, which regards all overlapping candidate frames as redundant. Soft-NMS believes that there may be effective detection targets even in candidate frames with more overlaps, but the larger the overlapping area, the lower the possibility. Therefore, in the algorithm execution, instead of directly deleting the frame whose IoU is greater than the threshold, the connection between IoU and the confidence score is first established. The formula is as follows, and the Gaussian penalty function is used to reduce the confidence score of the candidate frame until it is lower than the IoU threshold. Avoid deleting the correctly positioned candidate boxes in the case of overlapping detection targets, thereby improving the detection accuracy and recall rate of the model. In the formula, D is the set of all valid candidate frames retained. When IoU is 0, the confidence score remains unchanged. When 0<iou<1, the confidence score decays. Figure 3 displays the Soft-nms algorithm's pseudocode.

Algorithm 1 Soft-NMS

Input: $\mathrm{B} = \{b_1 \ldots b_N\}, \mathrm{S} = \{s_1 \ldots s_N\}, \mathrm{N_t}$
begin
  $\mathrm{D} \longleftarrow \{\}$
  while B≠**empty** **do**
    $\mathrm{m} \longleftarrow \mathrm{argmax}\, \mathrm{S}$
    $\mathrm{M} \longleftarrow \mathrm{b_m}$
    $\mathrm{D} \longleftarrow \mathrm{D} \cup \mathrm{M}; \mathrm{B} \longleftarrow \mathrm{B} - \mathrm{M}$
    for $\mathrm{b_i}$ **in** B **do**
      $\mathrm{s_i} \longleftarrow \mathrm{s_i f(iou(M, b_i))}$
    end
  end
  return D,S
end

Figure 3. Soft-nms algorithm pseudo code

2.5 Add detection layer

The initial Yolov5 model has three detection frames, and the output sizes are 80×80, 40×40, 20×20, which are used for the recognition of small, medium and large targets respectively. However, foreign objects on the transmission line are often of different scales, and The background is complex and easily occluded, and the shallow features contain more details and spatial information. Therefore, a shallow detection layer with a scale of 160×160 is added to increase the range of network detection scales as shown in Figure 4 and improve the model's the capacity to recognize objects of various sizes,detection performance.

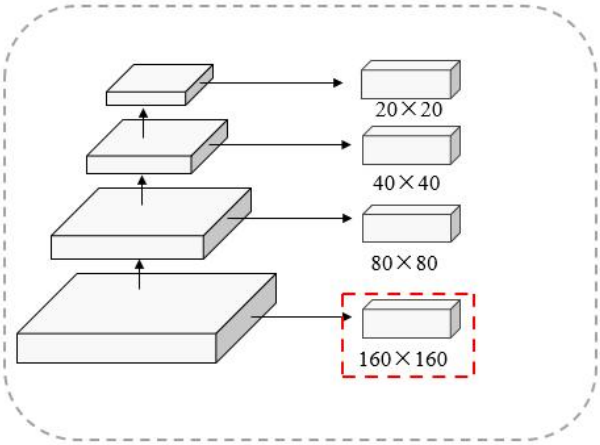

Figure 4. Add detection layer

3. EXPERIMENT

3.1 Dataset and model training

Since there is no public data set for foreign objects on transmission lines, the data sets in this experiment are 3726 private data sets constructed for three types of foreign objects: bird's nest, kite, and balloon. Since there aren't enough training samples, the data set needs to be expanded. In deep learning, training samples are expanded. Common methods for augmenting datasets include cropping, flipping and rotating, scaling, shifting, adding Gaussian noise and color dithering, etc. In this paper, Through the use of techniques like rotation, shifting, and cropping, the data set is increased to 6500, and the training set to the verification set to the test set ratio is 8:1:1.

If the model is trained from scratch, the effect will not be ideal, especially if the quantity of the training sample is insufficient. The ImageNet dataset has been widely used as a pre-training dataset because its data samples are large enough and more than 1,000 categories contained in the dataset come from daily life, making it suitable for training universal models. Therefore, in order to avoid overfitting due to insufficient training samples, this paper first pre-trains on the ImageNet dataset through transfer learning to obtain pre-training weights, so as to avoid the backbone feature network weights from being too random. First freeze the weights in the first 169 layers of the model backbone network, and train the remaining layers. At this time, the network for feature extraction won't change. Unfreeze the previously frozen layers for training, fine-tune the weight parameters of all layers, and finally the trained model can be used for foreign object recognition on iron towers. Since the learning rate and batch size of the freezing and unfreezing stages are different, as well as the optimizer used, Table 1 displays the experimental parameter.

Table 1. Experimental parameter

Experimental parameters				
Learning stage	Batch_size	Optimizer_type	Learning_rate	Epoch
Freeze	16	sgd	1×10^{-2}	50
UnFreeze	8	adam	1×10^{-4}	250

3.2 Experiment configuration and evaluation indicators

The experimental platform's hardware environment, as designed in this paper uses the CPU as Interl Core i7-11700K, the GPU as NVIDIA GeForce RTX3070TI, and the memory as 32GB. The deep learning framework is the foundation of the software environment pytorch1.8.1 and python3.9. To evaluate the model's effectiveness, different target detection algorithms are used on the same data set to compare with this paper's algorithm, and the accuracy, average precision, mAP and model detection speed FPS commonly used in target detection are used as important evaluation indicators of the model. P_{re} is used to measure the likelihood that all samples predicted by the model to be positive are in fact positive. Precision rate and recall rate added together form the area under the P-R curve or AP, which is used to measure the quality of the recognition precision of the model on a single category. mAP represents the average AP of non-category samples, and the number of frames the model recognizes per second is referred to as FPS. TP stands for the proportion of positive samples deemed positive, FP for the proportion of negative samples deemed positive, and FN for the proportion of positive samples deemed negative.

$$\mathrm{P}_{re} = \frac{TP}{TP + FP} \tag{4}$$

$$AP = \int_0^1 p(r)dr \tag{5}$$

3.3 Analysis of results

In order to gauge how the model's performance would be affected by the improvement measures described in this paper, the improved results are analyzed by means of ablation experiments. The test results are displayed in Table 2. The RepVGG module increased the mAP value from 76.5% to 77.2%, indicating that the introduction of the RepVGG module can improve the model's capacity to extract features.Compared with the second group of experiments and the third group of experiments, after adding the CBAM module, the mAP value increased by 1.2%, indicating that the model

combined The attention mechanism module can suppress the influence of other negative information and strengthen the network's capacity to express features. Comparing the third group of experiments with the fourth group of experiments, the mAP value increased to 79.1% after using Soft-NMS instead of the conventional NMS algorithm, indicating that Soft-NMS can improve the recall rate of the model when similar objects overlap. Comparing the fourth group of experiments with the fifth group of experiments, the mAP value increased from 79.1% to 80.6% once the detecting layer has been added. The results showed that the accuracy of the model's detection of foreign objects of different scales was improved and the rate of missed detection was reduced.

Table 2. Ablation Experiment Results

algorithm	Group1	Group2	Group3	Group4	Group5
Yolov5	√	√	√	√	√
RepVGG		√	√	√	√
CBAM			√	√	√
Soft-NMS				√	√
Detection layer					√
mAP/%	76.5	77.2	78.4	79.1	80.6

The technique in this work is contrasted with various popular target detection models that are currently in use, such as Faster-RCNN, SSD, Yolov4 and Yolov5, to verify the advantages of the improved Yolov5 model in the detection of foreign objects on transmission lines. The comparison figures are displayed in Table 3.The improved algorithm proposed in this paper has a mAP value of 80.6% in foreign object detection on transmission lines, which is 4.1%, 5.9%, 11%, and 5.5% higher than Yolov5, Yolov4, SSD, and Faster-RCNN respectively. The model of two-stage detection Faster-RCNN Although the detection accuracy is high, the size of the model is too large, which is not suitable for carrying in mobile devices. Compared with other single-stage models in terms of model size, the improved model has a large increase in size compared to the Yolov5 model by 17.7MB. reduce.

Table 3. Comparison of the results of several algorithms' recognition

algoritm	mAP%	Size/MB	AP/%		
			Nest	ballon	skite
Faster-RCNN	75.1	987.1	73.3	75.7	76.5
SSD	69.6	144.5	65.1	70.2	73.6
Yolov4	74.7	267.3	72.2	75.4	76.7
Yolov5	76.5	51.7	74.8	76.3	78.5
Improved-Yolov5	80.6	68.4	79.2	80.9	81.8

4. CONCLUSION

Given the complicated situation with the foreign objects on the transmission line, they are easily blocked by tower poles, resulting in low detection accuracy. This paper selects the Yolov5 algorithm as the baseline model and combines the RepVGG module, integrates the CBAM attention mechanism, increases the detection layer and optimizes NMS. Algorithms and other methods are integrated to improve the model’s ability to detect foreign objects on transmission lines.Experimental results show that compared with the original Yolov5, the detection accuracy of bird’s nests, balloons,

and kites that exist on transmission lines has respectively increased by 4.4%, 4.6%, 3.3%, and the mAP value increased by 4.1%. Compared to other widely-used target identification methods, it also has certain advantages. Therefore, this algorithm can realize the identification of common bird's nests, balloons, and kites on transmission lines. , which can meet the requirements of foreign object identification tasks in daily power inspection, and has certain reference significance.

REFERENCES

[1] Bharati, P., Pramanik, A., Deep learning techniques—R-CNN to mask R-CNN: a survey[J]. Computational Intelligence in Pattern Recognition, pp. 657-668 (2020).

[2] Ren, S., He, K., Girshick, R., et al., Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 28 (2015).

[3] Redmon, J., Divvala, S. K., Girshick, R. B., et al., You Only Look Once: Unified, Real-Time Object Detection [J]. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788) (2015).

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

[5] Bochkovskiy, A., Wang, C. Y., Liao, H. Y. M., Yolov4: Optimal speed and accuracy of object detection [J]. arXiv preprint arXiv:200410934 (2020).

[6] Wu, W., Liu, H., Li, L., et al., Application of local fully Convolutional Neural Network combined with YOLO v5 algorithm in small target detection of remote sensing image[J]. PloS one, 16(10): e0259283 (2021).

[7] Jeong, J., Park, H., Kwak, N., Enhancement of SSD by concatenating feature maps for object detection[J]. arXiv preprint arXiv:1705.09587 (2017).

[8] Ding, X., Zhang, X., Ma, N., et al., Repvgg: Making vgg-style convnets great again[C]. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp.13733-13742 (2021).

[9] Woo, S., Park, J., Lee, J. Y., et al., Cbam: Convolutional block attention module[C]. In Proceedings of the European conference on computer vision (ECCV). pp.3-19 (2018).

[10] Bodla, N., Singh, B., Chellappa, R., et al., Soft-NMS--improving object detection with one line of code[C]. In Proceedings of the IEEE international conference on computer vision, pp.5561-5569 (2017).