



ARTICLE

A Transmission and Transformation Fault Detection Algorithm Based on Improved YOLOv5

Xinliang Tang¹, Xiaotong Ru¹, Jingfang Su^{1,*} and Gabriel Adonis²

¹School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang, 050000, China

²Department of Computer Science and Information Systems, Birkbeck Institute for Data Analytics, London, B100AB WC1E 7HX, UK

*Corresponding Author: Jingfang Su. Email: sujingfang1980@hebust.edu.cn

Received: 04 January 2023 Accepted: 29 April 2023 Published: 08 October 2023

ABSTRACT

On the transmission line, the invasion of foreign objects such as kites, plastic bags, and balloons and the damage to electronic components are common transmission line faults. Detecting these faults is of great significance for the safe operation of power systems. Therefore, a YOLOv5 target detection method based on a deep convolution neural network is proposed. In this paper, Mobilenetv2 is used to replace Cross Stage Partial (CSP)-Darknet53 as the backbone. The structure uses depth-wise separable convolution to reduce the amount of calculation and parameters; improve the detection rate. At the same time, to compensate for the detection accuracy, the Squeeze-and-Excitation Networks (SENet) attention model is fused into the algorithm framework and a new detection scale suitable for small targets is added to improve the significance of the fault target area in the image. Collect pictures of foreign matters such as kites, plastic bags, balloons, and insulator defects of transmission lines, and sort them into a data set. The experimental results on datasets show that the mean Accuracy Precision (mAP) and recall rate of the algorithm can reach 92.1% and 92.4%, respectively. At the same time, by comparison, the detection accuracy of the proposed algorithm is higher than that of other methods.

KEYWORDS

Transmission line; YOLOv5; multi-scale integration; SENet

1 Introduction

The transmission line is the carrier of electric energy transmission, so it is very important to ensure the safe and stable operation of the transmission line. The construction of transmission lines in China (especially the construction of ultra-high voltage lines) has a very tight schedule [1], so it challenges the security construction of transmission lines. The transmission line is often affected by climate change and foreign matters (such as kites, plastic bags, balloons, etc.), especially when foreign matters are hanging on the transmission line, they will greatly shorten the discharge distance, cause harm to pedestrians and vehicles under the transmission line, and in serious cases, lead to power transmission interruption, resulting in a large area power outage [2]. Therefore, it is very necessary to detect



foreign matters on the transmission line and overhaul electronic components on the transmission line. Common foreign matters of the transmission line include bird nests, plastic garbage bags, kites, etc., as shown in Fig. 1.



Figure 1: Legend of fault type

However, with the continuous expansion of the scale of the transmission network, the method of relying on manual power transmission line fault detection has exposed the shortcomings of low efficiency and low safety factor. The emergence of intelligent inspection technology has gradually replaced the manual inspection method, liberated some manpower, and improved efficiency. Aerial photograph inspection based on aircraft is a commonly used intelligent inspection method for foreign matters in transmission lines [3]. This method judges whether the transmission line has a fault based on the depth learning algorithm by transmitting the captured transmission circuit information, which is of great significance for the accurate detection of transmission line faults.

The target detection method for intelligent patrol is mainly to extract the features of the target to be detected according to its characteristics, and then to segment and identify the target. Its detection accuracy is often difficult to guarantee. In 2006, the emergence of deep learning theory injected new vitality into the field of vision-based object detection [4]. The essence is to extract features from external input data by establishing more layers of neural networks, and then make decisions. After years of development, deep learning has made excellent achievements in the image. The network architecture based on Convolutional Neural Network (CNN) has shown good results in image classification, target detection, target tracking, semantic segmentation, instance segmentation, and other fields. At present, the target detection algorithm based on deep learning is mainly divided into the two-stage target detection algorithm and the single-stage target detection algorithm. R-CNN [5], Fast R-CNN [6], Faster R-CNN [7], and Cascade R-CNN [8] are typical two-stage target detection algorithms. These algorithms generate candidate regions first, and then classify and position regression. Therefore, although their accuracy is high, they are not fast enough. Therefore, Redmon et al. proposed one-stage algorithms such as YOLO [9], YOLOv2 [10], and YOLOv3 [11], which abandoned the candidate box generation stage and directly classified and regressed the target, thus improving the real-time detection

speed of the target detection algorithm. Among the YOLO series, YOLOv3 is the most classic. It adds a feature pyramid to YOLOv2, making its top-down multi-level prediction structure enhance the network's detection ability for small targets.

YOLOv5 is an algorithm widely used in industrial detection after YOLOv3. It can meet real-time requirements while maintaining high accuracy, and different detection models can be selected according to different working environments and detection tasks. Song et al. [12] applied YOLOv5 to the real-time detection of dense vehicles on roads. Sruthi et al. [13] carried YOLOv5 on the search and rescue UAV for human detection during the search and rescue process. Dong et al. [14] proposed a method for detecting pulmonary nodules by combining YOLOv5 with an attention mechanism. YOLOv5 uses the depth residual network to extract target features; and uses the Path Aggregation Network (PANet) structure to complete multi-scale prediction. However, when YOLOv5 extracts features to obtain the maximum feature map, it still conducts three down samples, and the target feature information is lost too much, so the detection of small targets is not ideal.

Therefore, considering the high requirements for detection accuracy and efficiency in the context of transmission and substation fault detection, this paper proposes an improved YOLOv5-based transmission and substation fault detection algorithm. Firstly, Mobilenetv2 is used to replace Darknet-53 as the backbone network for target feature extraction, which reduces the number of model parameters and improves the detection speed, as demonstrated in [15] and the experiments conducted in this paper. Secondly, the SENet attention model is introduced in the improved YOLOv5 network structure to enhance the saliency of fault targets in complex backgrounds, and new detection scales applicable to small targets are added for multiscale feature fusion, improving the overall detection accuracy of the detection network for transmission line faults.

2 Related Work

Power transmission and transformation fault detection are mainly to process data sets and train target detection models. Therefore, this paper investigates and summarizes the above two aspects of research work. Firstly, in terms of image processing, Huang et al. [16] proposed a foreign object detection method for power towers based on a coarse fine search strategy based on the analysis of the location and color characteristics of foreign bodies on a power tower. This method is based on the fact that foreign matters usually appear in the cross-arm area of the power tower pole. By roughly locating the high-voltage tower first, and then carefully searching for candidate areas of foreign matters, and screening out areas with high confidence based on the color statistical characteristics of foreign matters, the method eliminates most of the noise, effectively reduces missing and false detection, and improves the accuracy and efficiency of foreign matter detection. However, there is still a high false detection rate in the case of complex backgrounds. Ren et al. [17] proposed a method for enhancing the contrast of linear edges of transmission lines based on the principle of wavelet analysis, which can purposefully enhance the linear edges. Compared with the global image contrast enhancement method, it is more conducive to the edge extraction of transmission lines. Wang et al. [18] proposed a shape component recognition method based on local contour features. This method aims to identify the connecting components of vibration dampers and spacer rods installed on wires. The wire is segmented and analyzed to calculate the width variation and grayscale similarity of segmented wires, enabling the detection of broken strands and foreign object defects on the wire. Through testing on images of transmission lines captured by unmanned aerial vehicles during inspections, it has been demonstrated that this method can effectively detect broken strands and attached foreign object defects

on wires even under complex background conditions. All of the mentioned literature focuses on various image processing techniques employed in the target recognition process.

In the transmission line detection method based on deep learning, Zhou et al. [19] proposed the YOLOv3-Reparameterization Visual Geometry Group (Rep VGG) algorithm to improve the detection accuracy of foreign objects in transmission lines. The algorithm increases the number of RepVGGs and multi-scale target detection frames by modifying the backbone network. The experiment shows that the improved network can significantly improve the detection performance of small targets. However, the data set used is less, and there is no systematic foreign object classification on transmission lines for various scenarios. Therefore, it is necessary to add data sets of different scenarios and different angles to improve the detection capability of the model. Zou et al. [20] proposed an improved YOLO algorithm, which optimizes the BCA-YOLO network for small target detection, and CSP2_X replaced_X in YOLOv5 with CSP_CA, and added a layer of small target detection layer, and replace the Feature Pyramid Network (FPN) structure in the original network with BiFPN with less computation. The experimental results show that compared with the traditional YOLOv5 network mAP_{0.5} increased by 3.8%. Hao et al. [21] proposed a YOLOv3 transmission line fault detection method based on a convolutional block attention model. In the YOLOv3 algorithm framework, the convolutional block attention model is fused to enhance the saliency of the fault target area in the image. Then, the non-maximum suppression method was improved by introducing the Gaussian function to reduce the rate of missing detection of partially occluded targets. Li et al. [22] improved the YOLOv3 network, used Mobilenetv2 instead of Darknet-53 as the backbone, and used depth separable convolution to replace the 3×3 convolution kernel in the detection head. At the same time, the fully Convolution First Level Object Detection (FCOS) class encoding and decoding scheme are used to reduce the network complexity. Experiments show that the improved model has a smaller model size and higher detection speed than other existing models. Liu et al. [23] introduced an information aggregation algorithm based on YOLOx-S. This algorithm enhances relevant features and suppresses irrelevant features by aggregating spatial information and channel information in the feature map to improve detection accuracy. However, the detection accuracy of these methods is not high, which can not meet the balance between detection speed and accuracy.

This paper proposes a transmission line fault detection method based on improved YOLOv5, which solves the problem of low target detection accuracy caused by low fault target image saliency under complex backgrounds. By introducing the attention mechanism (SENet) and multi-scale feature fusion into the YOLOv5 algorithm, the saliency of the fault target in the complex background is enhanced, to improve the detection accuracy of the detection network. In addition, Mobilenetv2 is used as the backbone instead of CSP-Darknet53 in the network structure. This structure uses depth-wise separable evolution to reduce the amount of computation and parameters, and improve the detection rate.

3 Research Method

3.1 Target Detection Principle of YOLOv5 Algorithm

The structure of YOLOv5 [24] is shown in Fig. 2. The input end is enhanced with Mosaic data. Based on the initial anchor box, the prediction box is output based on the training data, and compared with the real box ground truth. The gap between the two is calculated, and then updated in reverse. At the same time, the original image is adaptively filled with black edges and scaled to a uniform size, and then sent to Backbone to improve the algorithm speed.

The backbone network mainly consists of the Focus structure and CSP structure. The Focus module performs the slicing operation on the pictures before they enter the backbone. First, the $640 \times 640 \times 3$ image is transformed into a 320×12 feature map, and then the 3×3 convolution operation is performed to form a $320 \times 320 \times 64$ feature map. The feature information of the original image is retained to the maximum extent while the downsampling. The CSP module divides the feature mapping of the base layer into two parts, and then merges them through a cross-phase hierarchy, reducing the amount of computation and improving the accuracy of the algorithm.

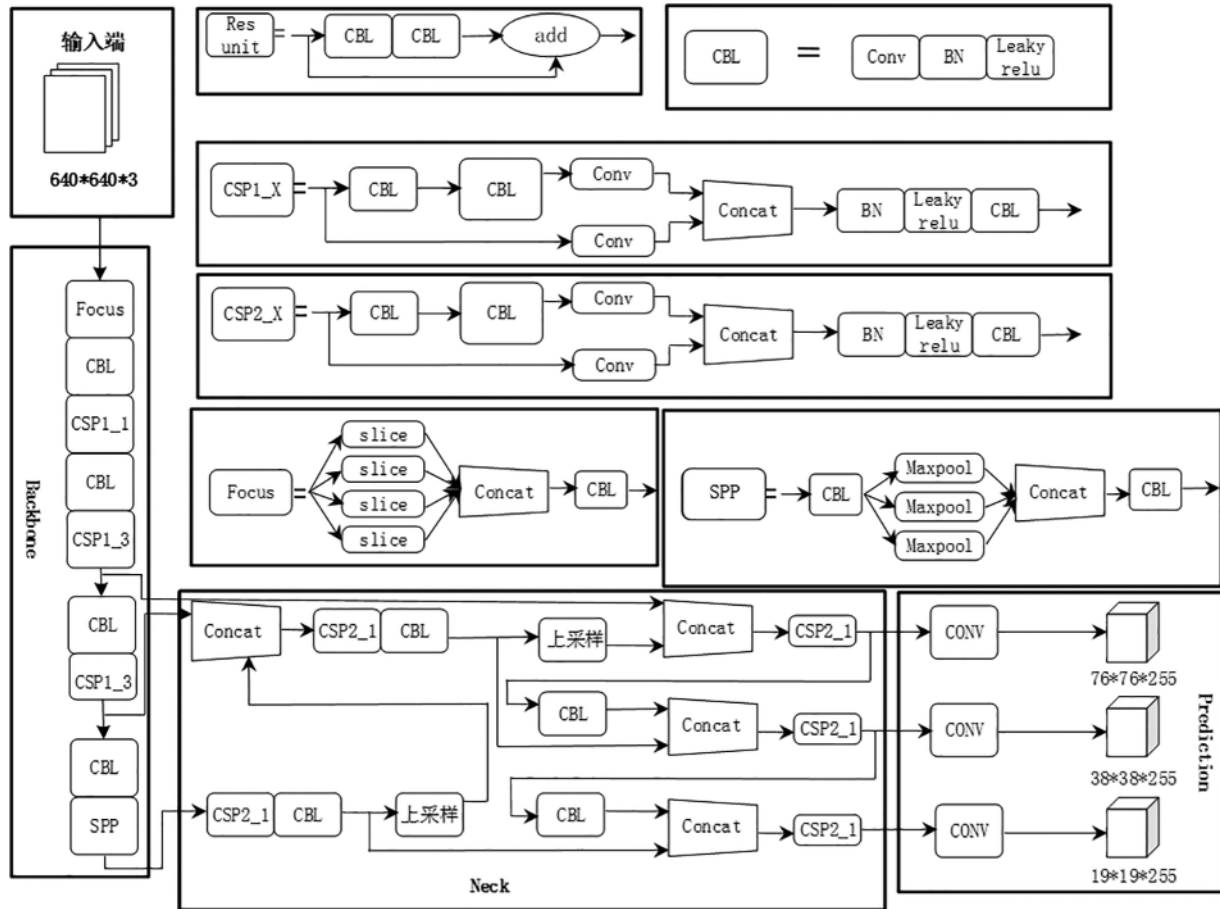


Figure 2: YOLOv5 network structure

The Neck structure of YOLOv5 adopts the structure of a multi-scale feature pyramid FPN + PANet. FPN enhances the semantic information of the feature map from top to bottom by upsampling and shallow feature fusion of the deep feature map, while PANet down samples the shallow feature map with strong location information from bottom to bottom with the deep feature map to enhance the feature fusion capability of the Neck network.

Finally, at the output end, YOLOv5 uses GIOU_Loss as the loss function of bounding box regression. It is defined as follows:

$$F_{Loss} = 1 - IOU + \frac{|C/(A \cup B)|}{|C|} \tag{1}$$

In formula (1), F_{Loss} represents the loss function, IOU represents the intersection and merger ratio, A represents the labeled box, B represents the prediction box, C represents the minimum bounding rectangle area of the prediction box and the labeled box, and AUB represents the union of the prediction box and the labeled box.

3.2 Headings Improved YOLOv5 Algorithm

3.2.1 Backbone Network

To improve the detection speed of the target detection network, this paper proposes a method, which uses Mobilenetv2 as the backbone network of YOLOv5. Replace the CSP-Darknet53 structure of YOLOv5 with the Mobilenetv2 structure. The detailed structure is shown in Table 1.

Table 1: Structure and parameters of mobilenetv2

Name	Structure
Conv	Conv321 BatchNorm ReLU6
Expanded_conv~expanded_conv_16	As shown in Fig. 3 Conv110
Conv_1	BatchNorm ReLU6

The Mobilenetv2 structure consists of a convolution layer (Conv), a batch normalization layer (BN), and a ReLU6 activation layer. The detailed structure of this module is shown in the figure. In Fig. 3, it consists of two point-by-point (PW) convolutions and one depth (DW) convolution. This module reduces the complexity of the convolution layer exponentially. Therefore, the YOLOv5 network effectively reduces the storage and computing overhead at the algorithm level, and improves the detection speed of foreign objects in the transmission line. Different from MobileNetv2 in the literature [25], the structure used in this paper removes the final average pooling layer and full connection layer. Because these two layers are mainly used for image classification, target detection is not required.

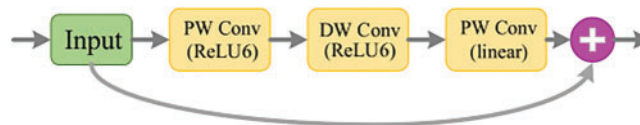


Figure 3: Schematic diagram of inverted residual module

3.2.2 Multi-Scale Feature Fusion

The foreign object detection of transmission lines based on the YOLOv5 network model is fast, but it is not effective in detecting small targets. In the YOLOv5 model, the original image is subsampled by 4 times, 8 times, and 16 times and sent to the feature fusion network and 13×13 Small size, 26×26

medium size, and 52×52 large size in the detection layer. The feature size is used to detect objects of different sizes. The original feature extraction model is shown in Fig. 4.

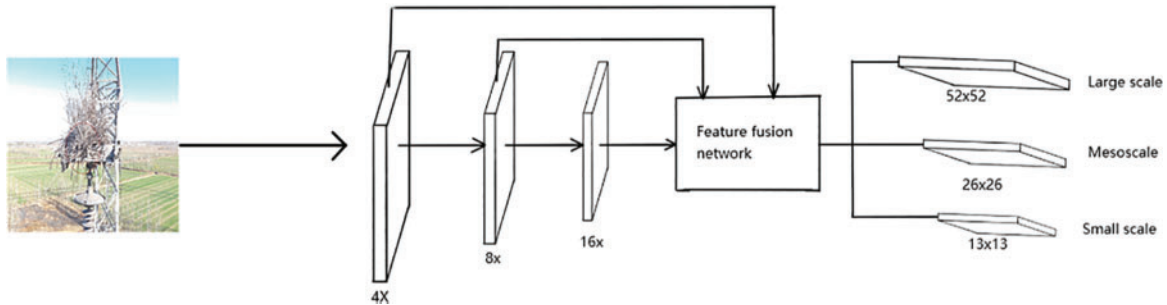


Figure 4: Original YOLOv5 feature extraction model

In the convolution neural network, the feature map obtained by deep convolution is rich in semantic information, but it will lose location information, and the location information of the target is not accurate enough. Therefore, adding down samp based onis of the original algorithm is conducive to small target detection. The original image is sent to the feature fusion network after 4 times, 8 times, 16 times, and 32 times down sampling $\times 13$ small sizes, 26×26 medium size, 52×52 feature maps of large size and new size, and feature maps of four sizes are used to detect targets of different sizes. The improved model is shown in Fig. 5. The improved new size feature map is conducive to better learning target features, making the detected target position more accurate, the receptive field of the feature map smaller, and improving the accuracy of small target detection.

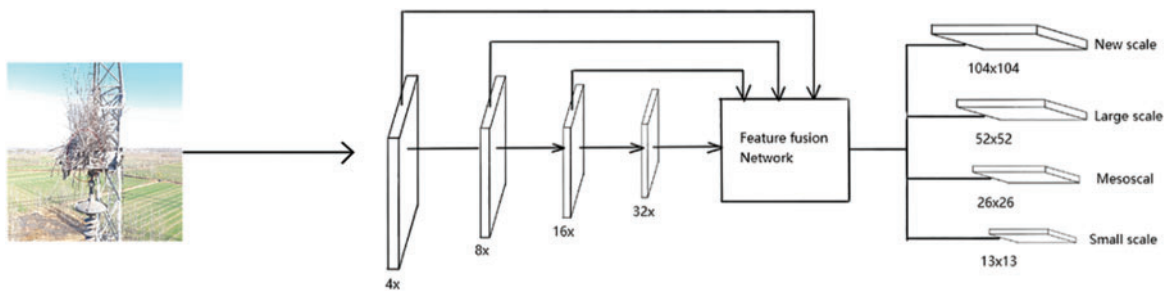


Figure 5: Improved feature extraction model

3.2.3 Integrate into SENet Attention Mechanism

Based on improving the structure of the YOLOv5 network model, the SENet module is added, which utilizes global average pooling to compress the feature map of the dataset image, reduce the feature dimension, then use Rectified Linear Unit (ReLU) function to activate it, and then use the full connection layer to return to the original dimension. In this way, the correlation of channel complexity can be well-fitted, and the number of parameters and calculations can be greatly reduced. Then a sigmoid function is used to obtain the normalized weight. Finally, scale operation is used to weigh the normalized weight of each channel, and the output and input features have the same number of weights. Fig. 6 shows the schematic diagram of the SENet module.

After the improvement, the main network of YOLOv5 was replaced by the Mobilenetv2 structure from the CSP-Darknet53 structure. Fig. 7 shows the Mobilenetv2 network structure on the right side, which simply includes point convolution, depth convolution, and point convolution. This main network focuses on extracting target object feature information, and the SENet module is inserted into

the gap between the second depth convolution and point convolution. The weight coefficients obtained by the deep convolutional feature data and the input attention mechanism are multiplied and fused, and finally input into the point convolution to obtain the final output. To verify the effectiveness of the proposed algorithm in two different scenarios, experimental results are compared and analyzed. Its fusion structure is shown on the left side of Fig. 7.

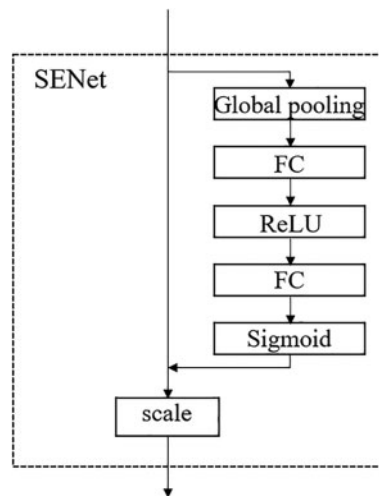


Figure 6: SENet module

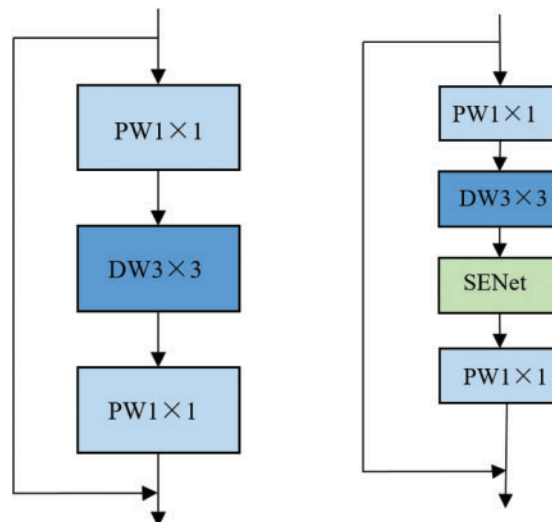


Figure 7: Result of integrating the SENet attention mechanism network

4 Experimental Results and Analysis

4.1 Data Set

In this paper, the data set was amplified by data enhancement. The data enhancement method is used to expand the original samples to form enhanced training samples, which is helpful to prevent over fitting and stopping learning features, and improve their generalization ability and robustness.

The amplified data set is divided into four categories, 2300 pieces in total. The data set is made into YOLO format, and 70% is randomly selected as the training set, 10% as the verification set, and 20% as the test set. Data enhancement refers to changing the image through transformation, which usually includes flipping, clipping, rotation, scaling and other operations. Fig. 8 shows the enhancement effect of some data sets.

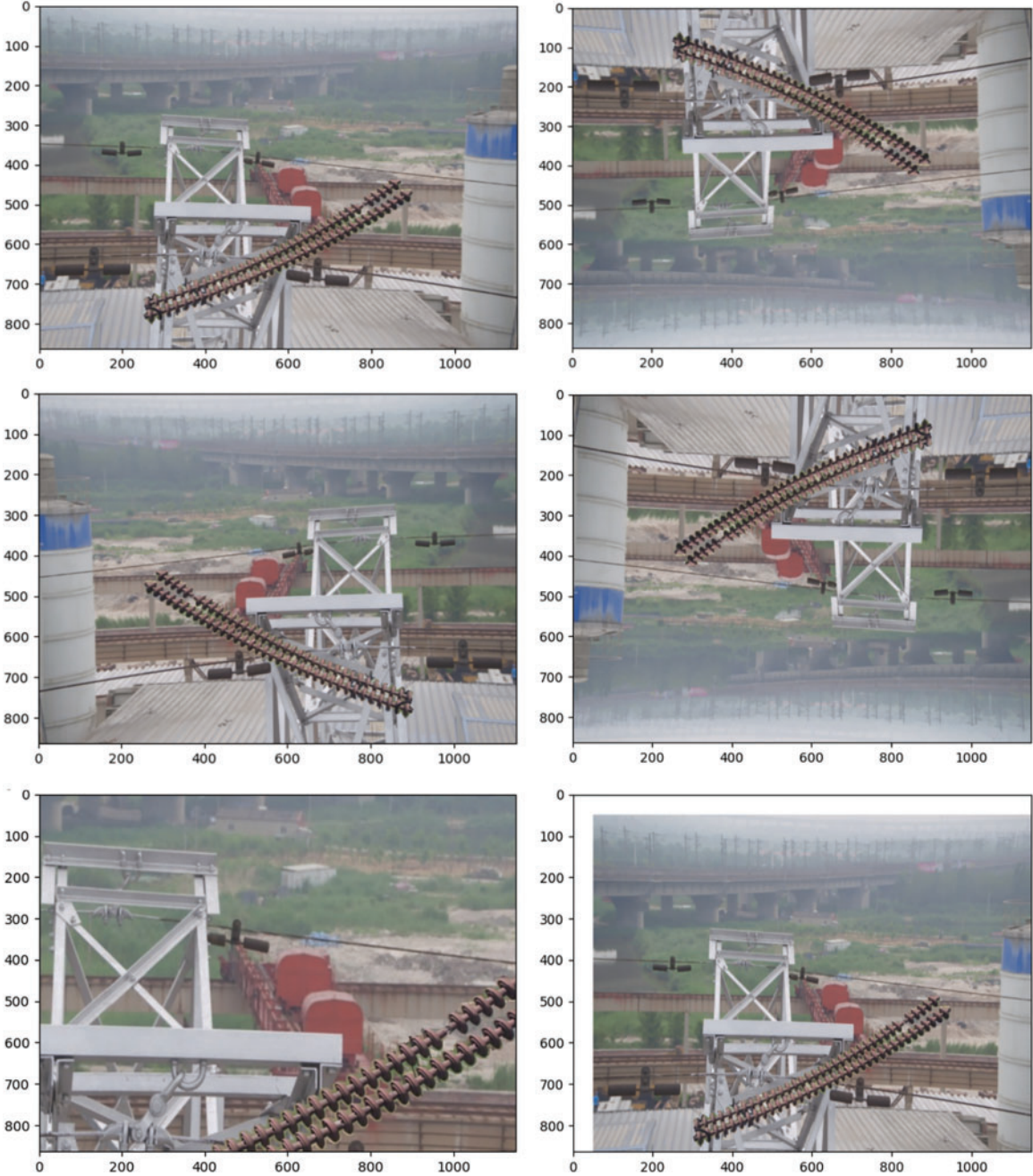


Figure 8: Data enhancement effect

4.2 Evaluating Indicator

To verify the performance of the model, this paper selects Precision (P), Recall (R), Average Precision (AP), and the mean Average Precision (mAP) to evaluate the detection performance of the model.

MAP needs to calculate P first. P represents the number of true positives (NTP) of a category C correctly recognized by the model in an image, accounting for the total number of targets identified by the model as category C (the sum of NTP and the number of false positives (NFP) of an image incorrectly identified as category C by the model). The accuracy rate P is:

$$p = \frac{NTP}{NTP + NFP} \quad (2)$$

AP is the average accuracy of one category of all images in the test set:

$$AP = \frac{\sum_{i=1}^n p_i}{n} \quad (3)$$

In formula (3): n is the number of test set pictures; I is the i-th image in the test set; P_i is the accuracy of a certain category of the i-th image.

MAP is the mean of the sum of the average accuracy of all categories of all images in the test set:

$$mAP = \frac{\sum_{k=1}^N AP_k}{N} \quad (4)$$

In formula (4): N is the number of target categories; K represents the kth category; AP_k represents the average accuracy of the kth category on all images in the test set.

4.3 Results and Analysis

Transmission line fault detection is based on deep learning, so it needs to configure good software and hardware environment. In terms of software, the operating system of Windows 10 is chosen in this paper, and the environment for the experiment is built in Python software. Python has a huge function library, which is convenient and fast to use in deep learning networks. After downloading and calling it directly, we use the improved YOLOv5 for the deep learning network model. The detection of foreign bodies in power lines requires deep learning network to train a large number of pictures, so we use NVIDIA GeForce RTX 3090 as a graphics accelerator to speed up the network detection.

In this experiment, we first compare and analyze the results of various indicators obtained by the algorithm before and after the improvement. To evaluate the improved model more objectively, we conduct ablation experiments, and use different algorithms to compare on the same dataset. The experimental results prove that the algorithm in this paper has significantly improved on different indicators.

The training data curve comparison results with the original YOLOv5 algorithm are shown in Fig. 9. It can be seen that the algorithm in this paper has a certain improvement in the three indicators of mAP, precision and recall: the precision reaches 92.7%, YOLOv5 is 85.4%, 7.3% higher. The average

accuracy mAP value in the figure is 92.1%, and the final recall rate is increased from 88.3% to 92.4%, an increase of 4.1%. The convergence speed of each index is better than the YOLOv5 algorithm. When the algorithm in this paper iterates to about 200 rounds, the indexes in Fig. 8 reach convergence. The improvement of each index proves the effectiveness of this algorithm.

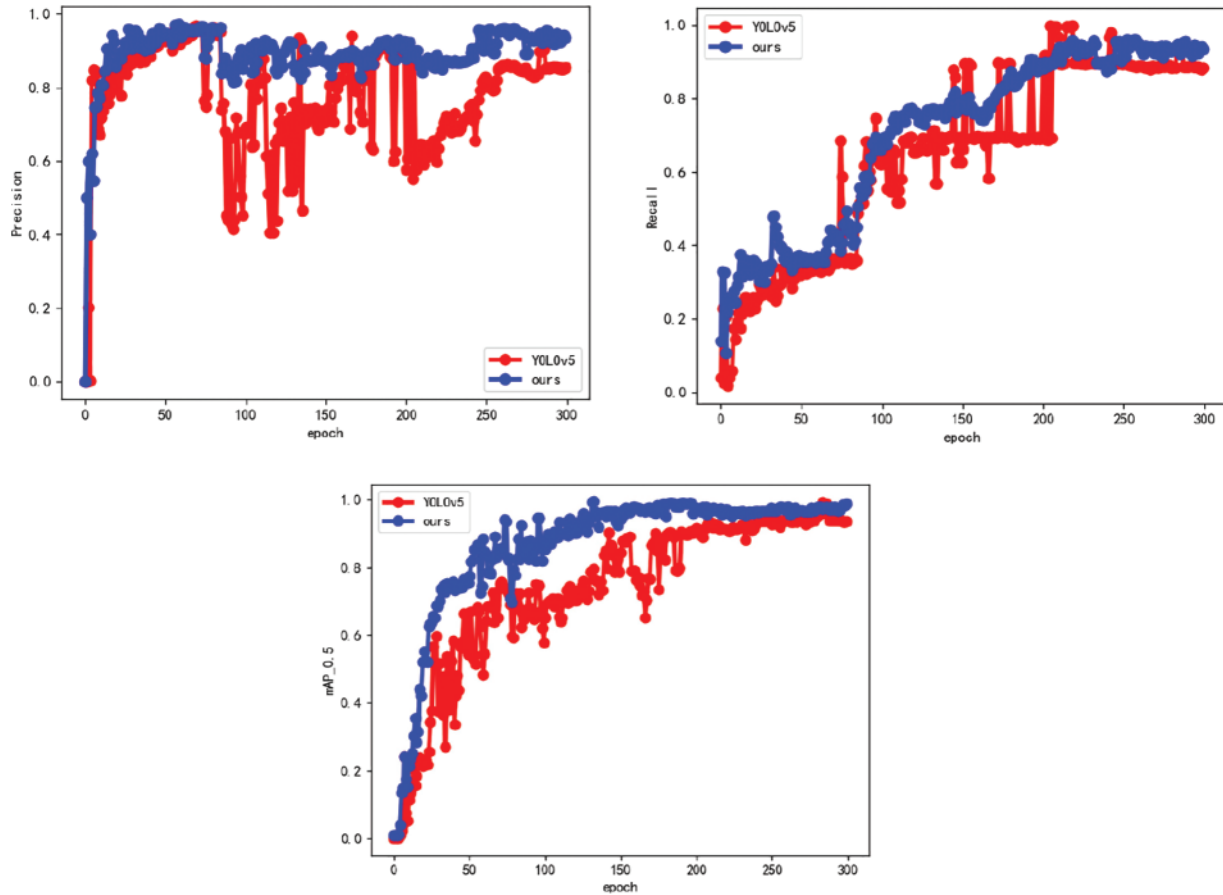


Figure 9: Comparison of mAP, precision and recall

In order to further study the impact of the improvement measures proposed in this algorithm on the model. We conducted ablation experiments, and the results are shown in Table 2.

“√” in the table indicates that this improved method has been added to the training. It can be seen from Table 2, the accuracy of this algorithm is 92.1%, which is significantly improved compared with 88.7% of the original YOLOv5 algorithm. And it can be seen that the algorithm after adding SENet is compared with the original YOLOv5. It can be seen that the accuracy of the algorithm after adding SENet is significantly improved by 3.8%, which proves the effectiveness of the SENet attention mechanism. In terms of detection speed, due to the added attention mechanism in the model, it is slightly slower, but the Mobilenetv2 structure just makes up for the deficiency and accelerates the detection speed.

Table 2: Impact of different method combinations on the model

SENet	Feature fusion	Mobilenetv2	mAP	Times/s
			88.7%	0.080
✓			90.5%	0.083
	✓		89.0%	0.082
		✓	86.5%	0.070
✓	✓		91.6%	0.085
✓		✓	90.1%	0.075
	✓	✓	88.7%	0.073
✓	✓	✓	92.1%	0.078

In order to illustrate the superiority of the algorithm in this paper compared with other target detection algorithms, the same data set is used to train on YOLOv3, Faster R-CNN, Single Shot MultiBox Detector (SSD) and other algorithm models commonly used for target detection. Finally, their results are summarized and compared, and the specific data are shown in Table 3. The average accuracy of this algorithm is much higher than that of the SSD algorithm. Although the precision of the Faster R-CNN model is good, the detection time is too long. The YOLOv3 algorithm is also inferior to YOLOv5 and the improved algorithm in this paper. Combined with the comparison between the previous algorithm and the original YOLOv5 algorithm, it can be seen that the algorithm in this paper is superior to other algorithms in the overall power transmission and transformation fault detection.

Table 3: Training results of different algorithms

Algorithm	mAP	Times/s
Faster R-CNN	88.3	2.875
SSD	85.2	0.068
YOLOv3	87.9	0.0514
YOLOv5	88.7	0.080
ours	92.1	0.078

In order to verify the algorithm in this paper, the scene where the target features are not obvious is selected for test verification, as shown in Fig. 10. The left side of each line of the graph is marked with a color box, and the right side is the test results of the algorithm in this paper. From the results in the graph, we can see that the improved algorithm in this paper can detect data with unobvious targets, but there are still undetected cases.



Figure 10: Inspection results

5 Conclusion

For power transmission and transformation fault detection, this paper proposes a fault detection method based on improved YOLOv5, which replaces the CSP-Darknet53 structure in the backbone network with Mobilenetv2, adds the SENet attention mechanism, and improves the network structure by adding a scale detection more suitable for small targets to the original scale detection. Through the comparison of experimental results, the improved algorithm has significantly improved the accuracy and detection speed. These two indicators are particularly important for the transformation of power transmission and transformation fault detection from manual patrol to robot operation, which improves the degree of grid intelligence.

Acknowledgement: Thanks to our team of mentors, our partner Professor Wang Zhenzhou and his company. They provide a good experimental environment and abundant engineering practice opportunities for the authors of this paper, provide inspiration for the research content and ideas of this paper, and provide us with the support of software and hardware resources.

Funding Statement: 1. Funding project: Key Project of Science and Technology Research in Colleges and Universities of Hebei Province. Project name: Millimeter Wave Radar-Based Anti-Omission Early Warning System for School Bus Personnel. Grant Number: ZD2020318, funded to author Tang XL. Sponser: Hebei Provincial Department of Education, URL: <http://jyt.hebei.gov.cn/>. 2. Funding project: Science and Technology Research Youth Fund Project of Hebei Province Universities. Project name: Research on Defect Detection and Engineering Vehicle Tracking System for Transmission Line Scenario. Grant Number: QN2023185, funded to W. JC, member of the mentor team. Sponser: Hebei Provincial Department of Education, URL: <http://jyt.hebei.gov.cn/>.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Xinliang Tang, Jingfang Su; data collection: Xiaotong Ru; analysis and interpretation of results: Xiaotong Ru, Gabriel Adonis; draft manuscript preparation: Jingfang Su, Xiaotong Ru. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The source code for the paper work is available at the following link: <https://github.com/18033754395/YOLOv5.git>. The data sets used in the work are not allowed to be disclosed due to the confidentiality of some projects involved.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] Z. Y. Liu, "Technology innovation of UHV AC transmission in China," *Power Grid Technology*, vol. 37, no. 3, pp. 567–574, 2013.
- [2] W. Z. Xing, X. B. Chen, Y. W. Xie and H. Y. Jia, "Research on Foreign body removal device of transmission line based on multi-rotor UAV and laser," *Electrical Technology*, vol. 468, no. 6, pp. 32–34, 2018.
- [3] X. P. Zhao, "Research on foreign body detection of transmission lines," Ph.D. Dissertation, Taiyuan University of Science and Technology, China, 2019.
- [4] S. W. Xu, C. M. Qiu, D. X. Zhang, X. He, L. Chu *et al.*, "Fault type identification of transmission lines based on deep learning," *Chinese Journal of Electrical Engineering*, vol. 39, no. 1, pp. 65–74, 2019.
- [5] R. B. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conf. on Computer Vision and Pattern Recognition*, Columbus, OH, USA, pp. 23–28, 2014.
- [6] R. Girshick, "Fast R-CNN," in *2015 IEEE Int. Conf. on Computer Vision (ICCV)*, Santiago, Chile, pp. 1440–1448, 2015.
- [7] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1137–1149, 2017. <https://ieeexplore.ieee.org/document/7485869>
- [8] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 6154–6162, 2018.
- [9] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection," in *2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 779–788, 2016.
- [10] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 6517–6525, 2017.
- [11] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," *arXiv:1804.02767*, 2018. <https://doi.org/10.48550/arXiv.1804.02767>

- [12] X. Song and W. Gu, "Multi-objective real-time vehicle detection method based on YOLOv5," in *2021 Int. Symp. on Artificial Intelligence and its Application on Media (ISAIAM)*, Xi'an, China, pp. 142–145, 2021.
- [13] M. S. Sruthi, M. J. Poovathingal, V. N. Nandana, S. Lakshmi, M. Samshad *et al.*, "YOLOv5 based open-source UAV for human detection during search and rescue (SAR)," in *2021 Int. Conf. on Advances in Computing and Communications (ICACC)*, Kochi, Kakkannad, India, pp. 1–6, 2021.
- [14] X. Dong, N. Xu, L. Zhang and Z. Jiang, "An improved YOLOv5 network for lung nodule detection," in *2021 Int. Conf. on Electronic Information Engineering and Computer Science (EIECS)*, Changchun, China, pp. 733–736, 2021.
- [15] Y. Xu, Y. Li, L. Zhang and B. B. Li, "Research on surface defect identification of aluminum strip by MobileNetV2 model incorporating coordinated attention," *Light Alloy Processing Technology*, vol. 51, no. 3, pp. 23–29, 2023.
- [16] L. L. Huang, "Research on intelligent Foreign body detection technology in transmission system," Ph.D. Dissertation, Shanghai Jiao Tong University, China, 2017.
- [17] G. X. Ren, "Research on intelligent video detection algorithm of transmission line icing and foreign body," Ph.D. Dissertation, Harbin University of Science and Technology, China, 2019.
- [18] W. G. Wang, J. J. Zhang, J. Han, L. Liu and M. W. Zhu, "Transmission line broken strands and foreign body defect detection method based on UAV image," *Computer Application*, vol. 35, no. 8, pp. 2404–2408, 2015.
- [19] H. Zhou, "Study on Foreign body detection method of power transmission line," Ph.D. Dissertation, Chongqing University of Technology, China, 2021.
- [20] H. J. Zou, L. B. Jiao, Z. J. Zhang, B. Y. Tang and Z. H. Liu, "An improved YOLO network for small target Foreign body detection in transmission lines," *Journal of Nanjing Institute of Technology (Natural Science Edition)*, vol. 20, no. 3, pp. 1672–2558, 2022.
- [21] S. Hao, R. Z. Ma, X. S. Zhao, B. Y. An, X., Zhang *et al.*, "Fault detection method of YOLOv3 transmission lines based on convolutional block attention model," *Power Grid Technology*, vol. 45, no. 8, pp. 2979–2987, 2020.
- [22] H. Li, L. Liu, J. Du, F. Jiang, F. Guo *et al.*, "An improved YOLOv3 for Foreign objects detection of transmission lines," *IEEE Access*, vol. 10, pp. 45620–45628, 2022. <https://ieeexplore.ieee.org/document/9763516/>
- [23] B. Liu, J. Huang, S. Lin, Y. Yang and Y. Qi, "Improved YOLOX-S abnormal condition detection for power transmission line corridors," in *2021 IEEE 3rd Int. Conf. on Power Data Science (ICPDS)*, Harbin, China, pp. 13–16, 2021.
- [24] H. F. Lv and H. C. Lu, "Research on traffic sign recognition technology based on YOLOv5 algorithm," *Journal of Electronic Measurement and Instrument*, vol. 35, no. 10, pp. 137–144, 2021.
- [25] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan *et al.*, "Feature pyramid networks for object detection," in *2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 936–944, 2017.