

Research Article

Open Access



An intelligent fault detection algorithm for power transmission lines based on multi-scale fusion

Tianyi Wu¹, Liming Wang¹, Xiangyi Xu¹, Lei Su¹, Wenjing He¹, Xinting Wang²

¹Shanghai Electric Power Research Institute, State Grid Shanghai Electric Power Company, Shanghai 200437, China.

²Shanghai Super High Voltage Electric Power Company, State Grid Shanghai Electric Power Company, Shanghai 200063, China.

Correspondence to: Tianyi Wu, Shanghai Electric Power Research Institute, State Grid Shanghai Electric Power Company, No. 171, Handan Road, Shanghai 200437, China. E-mail: wtyfenghua@163.com

How to cite this article: Wu, T.; Wang, L.; Xu, X.; Su, L.; He, W.; Wang, X. An intelligent fault detection algorithm for power transmission lines based on multi-scale fusion. *Intell. Robot.* **2025**, *5*(2), 474-87. <https://dx.doi.org/10.20517/ir.2025.24>

Received: 31 Mar 2025 **Revised:** 8 May 2025 **Accepted:** 9 May 2025 **Published:** 4 Jun 2025

Academic Editor: Simon Yang **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

With the rapid expansion of modern power grids, automated defect detection in high-voltage transmission lines has become a critical engineering challenge for preventing catastrophic failures and ensuring reliable electricity supply. While automated inspection has revolutionized power infrastructure maintenance, current vision-based methods still face three practical limitations in field applications: (1) susceptibility to complex background interference; (2) insufficient recognition accuracy for small-sized components; and (3) delayed response in real-time inspection scenarios. To address these industry pain points, this study develops a multi-scale fusion enhanced detection algorithm specifically optimized for power transmission line components. In response to these issues, this paper proposes an intelligent power transmission line defect detection algorithm based on multi-scale fusion, which introduce Coordinate Convolution, optimized decoupled detection head and improved loss function to solves the problems of low precision, poor robustness, and slow detection speeds faced by defect detection in power transmission network scenarios, laying a necessary theoretical foundation for subsequent practical applications.

Keywords: Power transmission line, multi-scale, intelligent algorithm, defect detection

1. INTRODUCTION

With the development of large-scale construction of the transmission network, the transmission tower and the power transmission line are a key part of the power system. Affected by the environment, natural aging



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



and other factors, the power transmission lines have a lack of pins, shockproof hammer off, nut rust and other defects. These defects will reduce the strength and service life of the power transmission line, which will cause economic losses to the power system and even serious accidents. Therefore, timely and accurate detection of defects in power transmission lines is of great significance for ensuring the safety of the power system^[1].

In recent years, scholars have put forward a variety of ideas and methods for power transmission line defect detection, including Ultrasonic Wave, X-ray detection based on manual detection^[2], infrared detection method^[3], laser scanning detection method, *etc.* These methods are non-automatic detection approaches, which require inspectors to observe through naked eyes, and long-term observation is prone to visual fatigue, false detection and missed detection. Traditional computer vision technology uses charged coupled device (CCD) cameras to collect image information of the object to be detected, and then analyzes it to obtain abnormal information, so as to obtain the corresponding defect information^[4] to complete defect detection.

Nowadays, drones have been widely applied to the inspection of power transmission lines, and algorithms such as deep learning have been extensively used in the defect recognition of images captured during drone-based power line inspections^[5–9]. Compared to defect identification by human eyes, machine vision technology has shown the advantages of visual, intuitive, efficient and fast defect detection tasks in power transmission line defect inspection scenarios.

You Only Look Once (YOLO) is a very famous, end to end real-time object detection algorithm by framing object detection as a single regression problem. Because of the remarkable speed and high accuracy of YOLO, many defect inspection algorithms, especially including power transmission line defect inspection, have drawn on the architectural framework of YOLO^[10–16].

The update from YOLOv1 to YOLOv3^[17,18] is mainly an improvement on the underlying framework structure: YOLOv1 uses a relatively simple convolutional neural network as a feature extractor, called Darknet. In the subsequent versions, YOLOv2 and YOLOv3 respectively use more advanced feature extraction networks, such as Darknet-19 and Darknet-53, which have better performance and expression capabilities, and in order to improve the detection ability of targets of different sizes, YOLOv2 and YOLOv3 introduce a multi-scale feature fusion mechanism. In addition, YOLOv2 and YOLOv3 introduce the concept of anchor boxes to provide candidate boxes of different scales. YOLOv12^[19] is the latest real-time object detector that redefines what is possible with faster detection speed, higher accuracy, and efficiency, and shines in the fields of industrial, medical, and security object detection.

In order to solve the problem of insufficient performance on edge computing devices, some researchers have proposed network structures with fewer parameters and lighter structures, such as Tiny YOLO and MobileNet YOLO. These networks achieve the goal of running in real time on low-power devices by reducing the number of network layers, employing lightweight convolution operations, or introducing low-resolution feature maps. They reduce the complexity of the model to some extent, but may sacrifice detection accuracy. In 2023, Liu *et al.* designed an edge-real-time object detector that can run in real time on edge devices to improve the inference speed without losing accuracy, and at the same time, designed an enhanced data augmentation method and a hybrid random loss function to improve the detection accuracy of small targets^[20]. Compared with the current mainstream YOLO series tiny models^[21], it shows good performance.

We find that the current transmission line defect detection methods do not effectively utilize edge information well, resulting in ignoring details. The contributions of our paper can be summarized as follows: (1) We introduced a residual structure and redesigned and optimized the decoupling head to make it more suitable for substation inspection scenarios; (2) By introducing coordinated coordinate convolution (CoordConv) to modify the feature extraction module, it provides more refined coordinate information for the subsequent convolutional layer, helps the network better understand the spatial relationship between the input data, and improves the accuracy and robustness of the model; (3) In terms of positioning loss, by introducing Shape Intersection over Union (Shape-IoU) and improving it to replace the original Generalized Intersection over Union (GIoU), the model considers the influence of the shape of the bounding box, the size of the bounding box and the distance between the prediction box and the real frame on the regression of the bounding box, so as to further improve the accuracy of loss calculation.

2. METHODS

Therefore, this paper uses the edge-real-time object detector as the basic model and improves and adjusts the shortcomings of the basic model, so as to further improve the detection accuracy while maintaining the original advantages.

2.1. Multi-scale fusion power transmission line defect detection method

The network structure of edge-real-time object detector mainly includes five parts: input, backbone network, feature fusion network, detection head, and output, as shown in [Figure 1](#).

The edge-real-time object detector proposes an improved method for image enhancement. First, multiple groups of images are processed using the Mosaic method, where the number of groups can be set according to the average number of labels in a single image in the dataset. The Mosaic method stitches multiple images together to produce a new composite image. Then, the last simple image enhancement data is mixed with the previous sets of Mosaic image enhancement data using the Mixup method. Finally, in the final output of the enhanced image, the bounding boxes of all original targets are still within the boundaries of the transformed enhanced image. By adopting the improved method in [Figure 2](#), we can avoid generating images without valid targets, improve the quality of the data by increasing the number of valid boxes, alleviate the negative impact of too few labels on training, and improve the performance of the detection model.

[Figure 3](#) shows the backbone network structure.

Firstly, four RepConv modules are passed through; the RepConv module comes from the Repvgg^[22] network, adopts the convolutional heavy parameterization method, and uses multiple branches and a single branch structure in the training and inference stages respectively to enrich the feature information and improve the computational efficiency. After four RepConv modules, the size of the feature map becomes $160 \times 160 \times 60$. This is followed by a feature extraction Module_1 module, as shown in [Figure 4](#). The Module_1 module consists of multiple RepConv and CBS modules. It effectively solves the problem of low detection accuracy of small targets in transmission line scenarios.

The CBS module is composed of convolution, batch normalization and SiLU activation functions, and the number of channels in the Edge_Module1 will become half of the original number of channels in the first two CBS modules, as shown in [Figure 5](#), in the subsequent modules, the number of input channels and output channels remains the same, and the output is the required number of channels after the last CBS module. Next, feature extraction will be carried out through three sets of MP_rep modules and

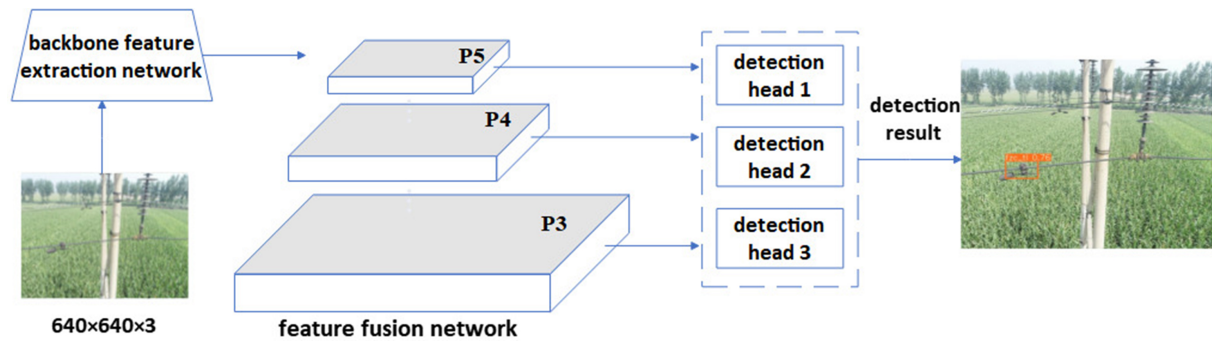


Figure 1. Overall structure of edge-real-time object detector.

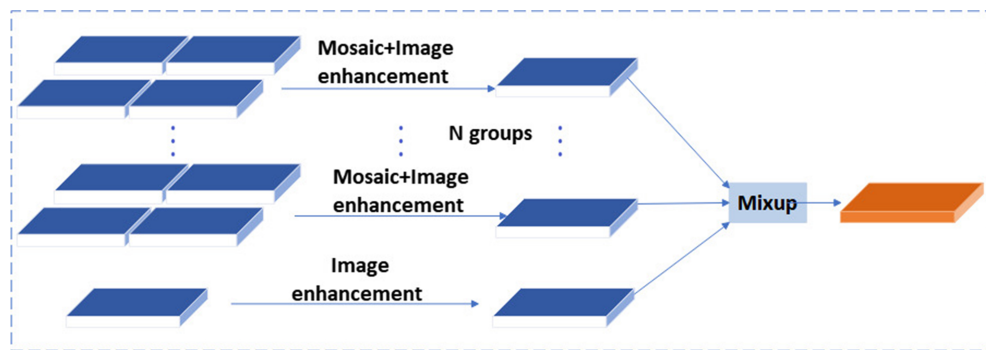


Figure 2. Improved data augmentation strategy.

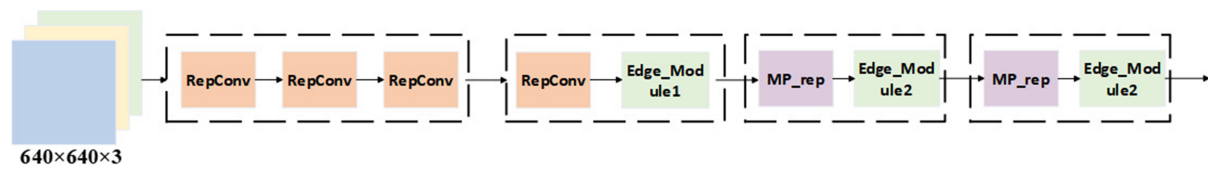


Figure 3. Backbone network.

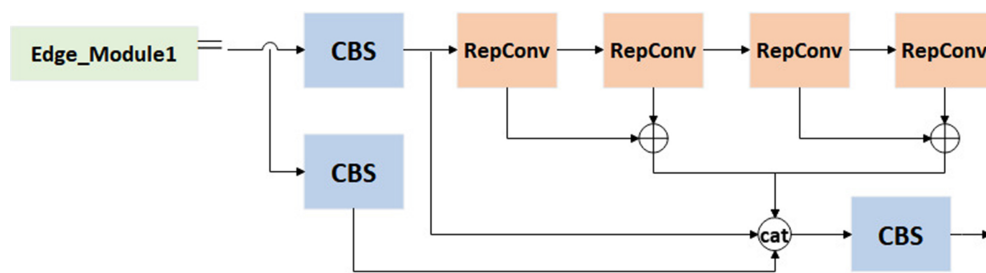


Figure 4. Edge_Module1 module structure.

Edge_Module2. The structure of the MP_rep module is shown in Figure 5, which mainly comprises Maxpool, CBS module and RepConv module, and the network structure of Edge_Module2 module and Edge_Module1 module is the same, as shown in Figure 6, except that the two RepConv modules are replaced by DepthwiseConv.

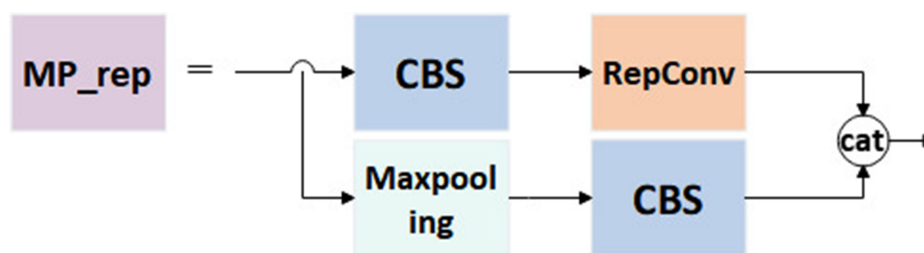


Figure 5. Mp_rep module structure.

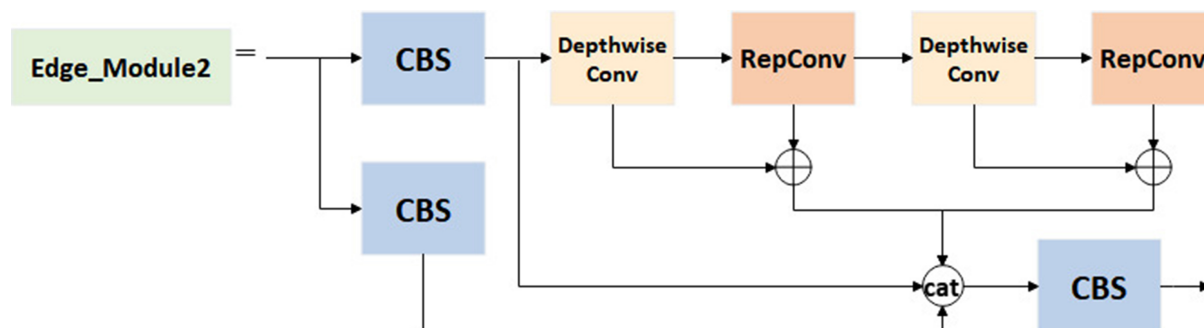


Figure 6. Edge_Module2 modular structure.

The feature fusion network utilizes the SPPCSPC module from YOLOv7 and the path aggregation network (PANet) to complete feature fusion at different scales, as shown in Figure 7. The SPPCSPC module uses four maximum pooling layers (1×1 , 5×5 , 9×9 , and 13×13) at different scales to pool the feature map, then combines these layers to provide the model with different receptive fields. This enables better differentiation between small and large targets, with the network structure illustrated in Figure 8. PANet’s core idea is to improve image comprehension by fusing features from different levels step by step. It consists of a top-down upsampling path to extract fine-grained features from the high-level feature map, and a bottom-up downsampling path to capture more global, semantically rich features from the lower-level feature map. The fusion of these two paths enables PANet to obtain rich multi-scale feature information and improve the accuracy of network prediction.

The P3, P4, and P5 outputs by the feature fusion network are transferred to the prediction network for prediction after the number of channels is adjusted by the RepConv module. Edge-real-time object detector takes a decoupling head as a prediction network and modifies it to make its inference faster. The decoupling head was first proposed in fully convolutional one-stage object detection (FCOS)^[23], and its main idea is to decompose the object detection task into two subtasks: classification and localization. Specifically, the decoupling head contains two separate branches: one for predicting the target’s category, and the other for predicting the target’s location. By decoupling the head, each subtask can learn and optimize independently, so as to better adapt to the characteristics of different classes and different scale targets, and improve the accuracy and robustness of detection. Edge-real-time object detector streamlines and optimizes the decoupling head, reducing the complexity of the model and improving the computational efficiency, as shown in Figure 9. The introduction of an implicit representation layer behind the convolutional layer can better capture the features of the target, and at the same time, reduce the cost and improve the regression performance in the inference process.

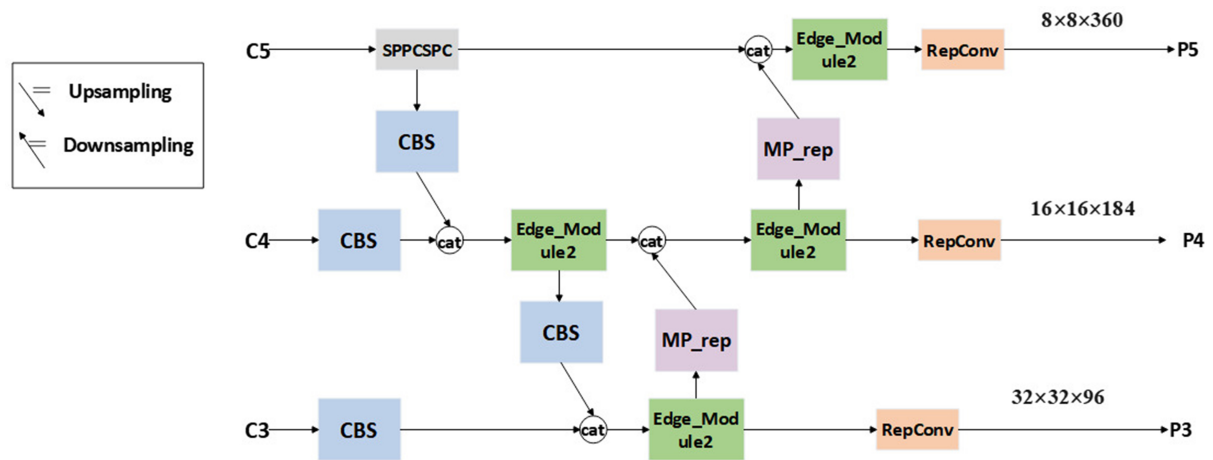


Figure 7. Feature convergence network.

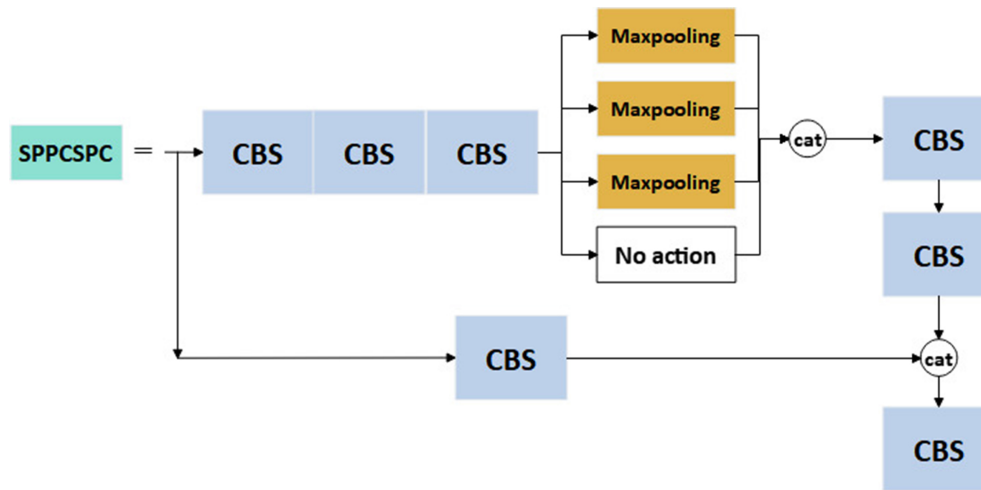


Figure 8. SPPCSPC network structure.

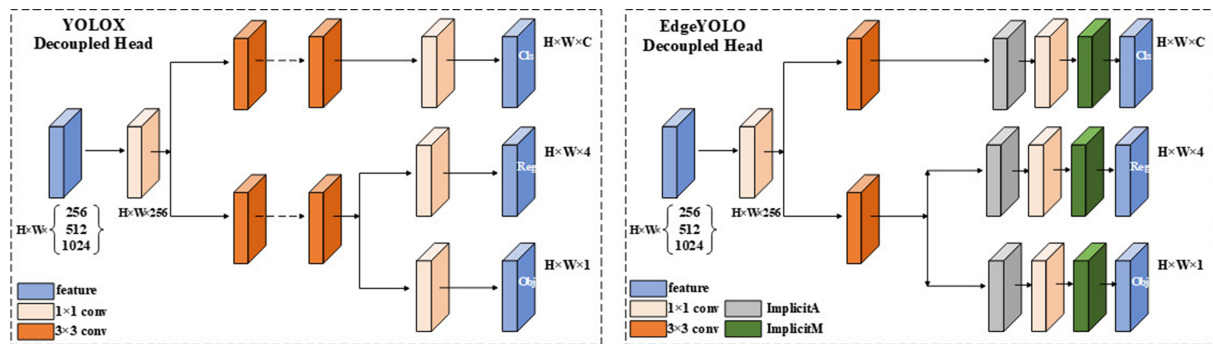


Figure 9. Structure diagram of the decoupling head.

2.2. Defect detection algorithm based on improved network

High-voltage environments, such as power transmission networks, often have complex light conditions and noise interference, which can affect the accuracy of machine vision inspections. At the same time, the defect

itself is a small target, which is difficult to detect, and there are problems with the detection accuracy. In order to solve these problems, based on edge-real-time object detector, this paper further improves the accuracy and stability of detecting power transmission line defects without increasing the number of additional parameters and memory overhead by modifying the feature extraction module in the backbone network, customizing the optimization decoupling detection head, and modifying the loss function.

2.3. Improved feature extraction module based on CoordConv

In order to solve the problem that the detection accuracy of small targets such as pins and nuts is not high in the complex scene of the transmission network, in the backbone network, the Edge_Module1 of the feature extraction module is modified by introducing CoordConv, which provides more refined coordinate information for the subsequent convolutional layer, helps the network better understand the spatial relationship between the input data, and improves the accuracy and robustness of the model.

In view of the problem of low detection accuracy for some small targets in power transmission line scenarios, this paper improves the Edge_Module1 of the backbone feature extraction module, and introduces the CoordConv^[24] in the Edge_Module1 module to replace the original ordinary convolution, and names it Module_CD, as shown in Figure 10.

CoordConv is a convolutional layer that introduces coordinate information, which can fuse position information into the convolution process without adding additional parameters and computational effort, which is an extension of ordinary convolution, as shown in Figure 11. It provides finer coordinate information for the subsequent convolutional layer (that is, two additional I and J channels are connected to the input channel to provide the filter), and establishes the mapping relationship between pixel space and Cartesian space to help the network better understand the spatial relationship between the input data, enhance the convolutional neural network's perception of position information, and improve the detection and recognition effect of small targets.

2.4. Custom decoupling head based on residual structure

It is necessary to address the deployment issues of large-scale networks on embedded devices. To meet the demand for real-time detection, it is necessary to perform lightweight processing on the basic detection model. In response to this issue, this paper introduces a residual structure and redesigns the decoupling head to make it more suitable for detection scenarios as shown in Figure 12. Through such optimization, we can not only improve detection speed but also maintain high detection accuracy.

To enhance the network's representational power, we first use the CoordConv module to uniformly adjust the number of channels in the three feature maps output by the feature fusion network to 256. Then, we add two residual structure branches to optimize gradient propagation and accelerate model convergence. By using skip connections to sum the inputs and outputs, the network can learn the residual parts, which can help the model better adapt to changes in input data, reduce dependence on precise predictions, and better cope with difficulties in detection caused by factors such as occlusion and image blur, where edges or details cannot be clearly displayed. This improves the model's robustness. With these improvements, we can achieve fast and accurate real-time detection of large-scale network models on embedded devices.

2.5. Improved loss function based on Shape-IoU

In the complex scenario of power transmission lines, due to the small size of the defects, even minor errors can lead to significant changes in Intersection over Union (IoU) values. Therefore, it is necessary to further optimize the localization loss to improve the accuracy of loss calculation and reduce the false positive and false negative rates at various scales. Research has found that the shape of small object ground truth (GT)

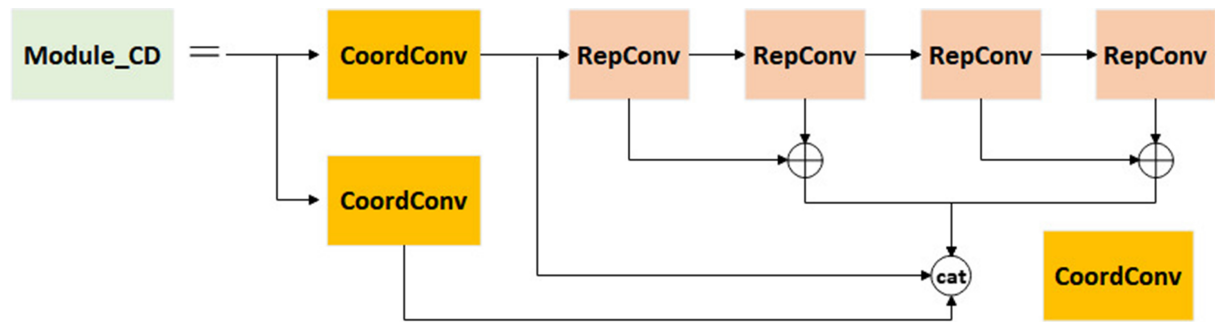


Figure 10. Module_CD structure diagram.

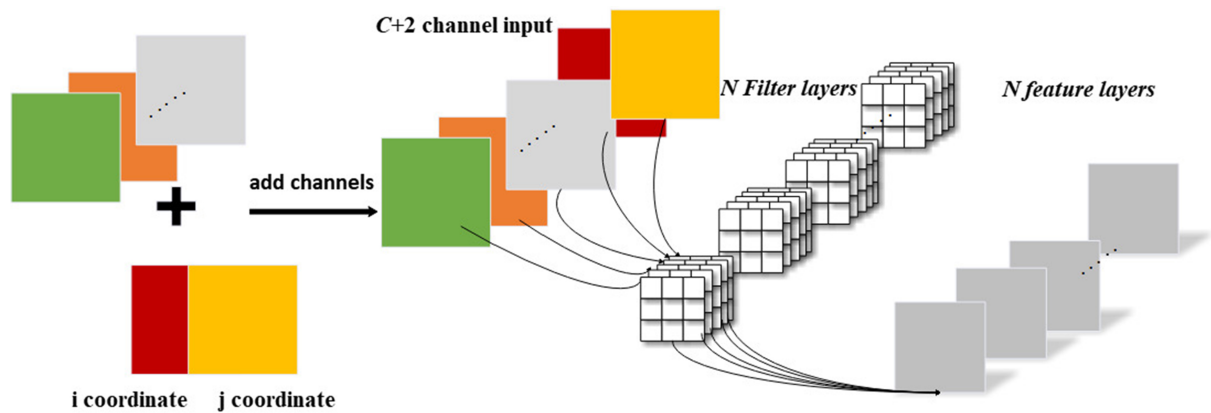


Figure 11. Coordinated coordinate convolution.

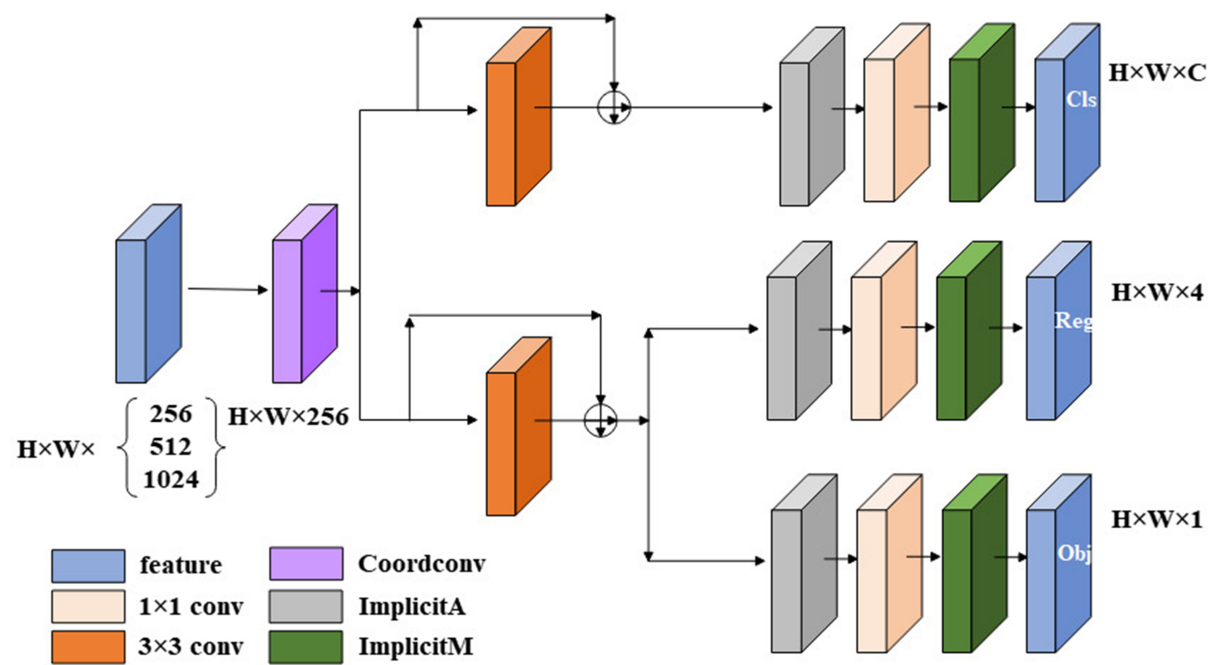


Figure 12. Custom decoupling headers.

boxes has a significant impact on the IoU value of bounding boxes, because even if there is some overlap between the bounding box and the GT box, the IoU value may still be low if the shapes do not match. Therefore, Shape-IoU^[25] is introduced and improved to make the model consider the shape of the bounding box itself, the size of the bounding box itself, and the impact of the distance between the predicted box and the GT box on bounding box regression simultaneously.

$$IoU = \frac{|b \cap b^{gt}|}{|b \cup b^{gt}|} \quad (1)$$

$$ww = \frac{2 \times (w^{gt})^{scale}}{(w^{gt})^{scale} + (h^{gt})^{scale}} \quad (2)$$

$$hh = \frac{2 \times (h^{gt})^{scale}}{(w^{gt})^{scale} + (h^{gt})^{scale}} \quad (3)$$

$$distance^{shape} = hh \times (x_c - x_c^{gt})^2 / c^2 + ww \times (y_c - y_c^{gt})^2 / c^2 \quad (4)$$

$$\Omega^{shape} = \sum_{t=w,h} (1 - e^{-\omega_t})^\theta, \theta = 4 \quad (5)$$

$$\begin{cases} \omega_w = hh \times \frac{|w - w^{gt}|}{\max(w, w^{gt})} \\ \omega_h = ww \times \frac{|h - h^{gt}|}{\max(h, h^{gt})} \end{cases} \quad (6)$$

$$L_{Shape-IoU} = 1 - IoU + distance^{shape} + 0.5 \times \Omega^{shape} \quad (7)$$

Where w^{gt} and h^{gt} denote the width and height of GT box, and w and h denote the width and height of anchor box. c is the diagonal distance of the minimum enclosing bounding box between b and b^{gt} . x_c, y_c are the coordinates of the center in anchor box. x_c^{gt}, y_c^{gt} are the coordinates of the center in GT box.

Due to the small size of bounding boxes for small objects, errors are likely to occur. To better reflect the overlap between the predicted box and the GT box, this paper adds the minimum enclosing shape C between the GT box and the predicted box to the Shape-IoU as the final localization loss function, as given in

$$L_{IoU} = 1 - IoU + 0.5 \times distance^{shape} + 0.5 \times \Omega^{shape} + 0.5 \times \frac{|C - B \cap B^{gt}|}{|C|} \quad (8)$$

Where C represents the smallest enclosing box between the true box and the predicted box, while B and B^{gt} denote the predicted box and the true box, respectively. This allows for a more accurate measurement of the matching degree between bounding boxes, improving the accuracy and robustness of the model.

3. RESULTS

3.1. Experimental environment and dataset

The training of an object detection model typically requires a certain amount of data to learn and generalize various patterns and features. Considering the particularity of the power transmission line scenario and the lack of publicly available professional datasets, this paper adopts a combination of online collection and self-production to construct the defect dataset. The whole dataset sample mainly includes three categories: nut rust, missing pins, and missing shockproof hammers. In this paper, these samples are manually annotated, and accurate object detection bounding boxes and category labels are provided for each sample. Then, the dataset images were enhanced with data enhancement processing, such as image cropping, rotation, noise

addition, and contrast processing, to enrich the diversity of the dataset. Through a well-designed data enrichment approach, the dataset can cover different situations and scenarios for model training, validation, and testing. The expanded dataset are first divided into training data and test sets according to the ratio of 9:1, and then subdivided into training sets and validation sets according to the ratio of 9:1 in the training data, which can ensure that we have enough data to train and verify the performance of the model, and can also set aside a part of the data as a test set to evaluate the performance of the model on unseen data.

3.1.1. Training parameters

We set 300 epochs in total with 5 warmup_epochs and 15 close_mosaic_epochs. The batch size is 8. Learning rate per image is 0.00015625. We use stochastic gradient descent (SGD) as an optimizer.

3.2. Evaluation indicators

For the defect detection task of transmission network, the average accuracy (AP), parameter quantity (Params) and frames per second (FPS) are used as the evaluation indicators of algorithm performance.

Average precision AP is a commonly used evaluation index in the field of object detection, which is used to measure the accuracy of the model in different types of object detection tasks. The higher the AP value, the better the performance of the model in the corresponding category of object detection tasks. When calculating the AP, the confidence score of each prediction box and the corresponding true label are determined based on the prediction results and the true label. Then, all prediction boxes are sorted according to the confidence score, and the prediction boxes are divided into positive or negative samples according to the set IoU threshold. This involves the calculation of precision and recall. Precision refers to the proportion of boxes that are correctly predicted to be positive out of all the boxes predicted to be positive, and Recall is the proportion of boxes that are correctly predicted to be positive among all true positive samples, as defined in

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

The AP is calculated by computing the area under the Precision-Recall curve.

$$AP = \int_0^1 p(r)dr \quad (11)$$

The number of parameters (Params) of a model refers to the total number of parameters in the training process, including weights, bias values, etc. The number of parameters is an important metric that can help us evaluate the complexity and scale of the object detection model and make sound decisions when designing and selecting the model.

FPS indicates the number of image frames processed within a unit of time. A higher FPS indicates that the algorithm can process images more quickly, thereby providing more real-time object detection results.

3.3. Ablation experiment

In order to verify the effect of the proposed algorithm on the improvement of detection performance, an ablation experiment was carried out. With the addition of Module_CD, a custom decoupling head, and an

improved loss function, four different models were obtained, denoted as A, B, C, and D. In this paper, these four models are tested on the above-mentioned self-made dataset to evaluate their detection accuracy and speed, and the experimental results are shown in [Table 1](#).

From [Table 1](#), it can be seen that as the base model A incorporates new modules, its detection accuracy gradually improves. Model B shows that after adding Module_CD, the model's accuracy increased by 0.3%, with only a 0.01M increase in parameters; however, the FPS decreased by 20 frames. This indicates that the introduction of the coordinated convolution feature extraction module indeed enhances the convolutional neural network's ability to perceive positional information, but the detection speed has decreased. After adding the custom decoupling head, model C achieved a 1.5% increase in AP compared to model B, with no change in the number of parameters, while the FPS significantly improved. This demonstrates that the custom decoupling head not only enhances detection accuracy but also improves detection speed. After adding the modified Shape-IoU on top of model C, the model's accuracy increased by another 0.7%. Although the FPS decreased by two frames, it maintained a balance between detection accuracy and processing speed.

To verify the superiority of the algorithm presented in this paper for power line defect detection, it was compared with current mainstream object detection algorithms on the self-made dataset mentioned above. The experimental results are shown in [Table 2](#).

From [Table 2](#), it can be seen that compared with RetinaNet and Faster-RCNN, the detection accuracy and model parameter count of the algorithm presented in this paper have been significantly improved. Compared to EfficientDet, the detection accuracy has increased by 2.6%, and the model parameter count has been reduced by 15.24 M. Compared with Single Shot MultiBox Detector (SSD), the improved model has increased detection accuracy by 0.4% while also reducing the model parameter count by 3.08M. When compared with YOLOv8-S, the improved algorithm maintains considerable detection accuracy while optimizing the model parameter count by 47.8%, greatly enhancing detection efficiency and better adapting to real-time detection tasks in power transmission lines. When compared with YOLOv11-n and YOLOv11-s, the detection accuracy has increased by 1.6% and 0.9%. When compared with YOLOv12-n and YOLOv12-s, the detection accuracy has increased by 1.4% and 0.7%. The number of parameters is only 5.82M, significantly less than YOLOv11-s (9.4M) and YOLOv12-s (9.3M), and even close to the scale of the ultra lightweight model YOLOv11-n (2.6M). Although the FPS metric is not as good as the YOLOv11 and YOLOv12 at similar parameter levels, our model has achieved a slightly lower frame rate in exchange for a higher AP_{50} . Compared with EdgeYOLO, the detection accuracy has increased by 4.0%. In summary, the improved algorithm in this chapter can better balance model lightweight and detection performance, with a comprehensive performance superior to other models, meeting the requirements for lightweight design and high precision. Even on low-cost devices with limited computing resources, the algorithm presented in this paper can achieve efficient defect detection.

3.4. Visualized comparative display

In order to verify the effectiveness of the improved algorithm, the above experiments are quantitatively analyzed, in order to show the detection effect of the improved algorithm more intuitively, this paper selects some difficult practical application scenario images for visual display, uses different color detection frames to select different types of targets to be detected from the image frames, and gives the corresponding labels and their confidence levels, as shown in [Figure 13](#).

Table 1. Ablation test results

Model	AP _{val}	AP _s	AP _M	AP _L	Parameter numbers	FPS
A	56%	25.3%	51.4%	59.6%	5.81M	57
B	56.2%	26.6%	49.6%	63.5%	5.82M	37
C	58.6%	30.1%	57.8%	65.8%	5.82M	57
D	60.0%	33.3%	58.0%	67.3%	5.82M	54

Table 2. Contrast test results

Model	Size	AP _{val}	AP ₅₀	Parameter numbers	FPS
Retinanet	640 × 640	46%	71.9%	37.97M	38
Faster-RCNN	640 × 640	40.3%	73.2%	136.73M	19
EfficientDet	640 × 640	57.4%	90.8%	21.06M	48
SSD	640 × 640	59.6%	91.9%	8.9M	31
YOLOV8-S	640 × 640	68.4%	93.3%	11.14M	50
YOLOV11-n	640 × 640	61.2	92.2%	2.6M	127
YOLOV11-s	640 × 640	65.5	92.9%	9.4M	101
YOLOV12-n	640 × 640	62.1	92.4%	2.6M	115
YOLOV12-s	640 × 640	64.9	93.1%	9.3M	96
EdgeYOLO	640 × 640	56%	92.8%	5.81M	57
This paper	640 × 640	60%	93.8%	5.82M	54

AP: Average accuracy; FPS: frames per second; SSD:

**Figure 13.** Comparison of detection results. (A) Our model; (B) YOLOV12-n.

4. DISCUSSION

In response to the complex background of power transmission lines and the serious issues of false positives and omissions for small targets, a new algorithm for power line defect detection has been proposed. The algorithm introduces coordinate convolution to reconstruct the feature extraction module, enhancing the backbone network's ability to capture key features. By optimizing the decoupled head, the model achieves faster detection while improving robustness against occlusion and image blur. Additionally, an improved Shape-IoU loss function enhances detection accuracy and stability without increasing model parameters or memory consumption, accelerating convergence. Experimental results show that the proposed algorithm achieves a superior balance between lightweight efficiency and detection performance, outperforming mainstream models and meeting the high-precision, low-computation requirements of power transmission line scenarios.

Moving forward, future research can combine attention mechanism and convolution operation to enhance algorithm capability, focusing on optimizing multimodal data fusion and computational efficiency while

improving system adaptability to complex environments such as low visibility and multipath interference. The integration of emerging sensor technologies and learning-based optimization strategies can further enhance system performance and generalization. This research bridges the gap between academic studies and industrial deployment, fostering intelligent inspection systems that align with the evolving demands of smart grid infrastructure and sustainable energy management.

DECLARATIONS

Authors' contributions

Made substantial contributions to conception and design of the study and performed data analysis and interpretation: Wu, T.; Wang, L.; Xu, X.; Su, L.; He, W.

Performed data acquisition and provided administrative, technical, and material support: Wang, X.

Availability of data and materials

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Financial support and sponsorship

This work was supported by the science and technology project of State Grid Corporation of China (No. 52094023002T).

Conflicts of interest

All authors are affiliated with State Grid Shanghai Electric Power Company, and they have declared that they have no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2025.

REFERENCES

1. Zuo, F.; Liu, J.; Fu, M.; Lu, J.; Liu, H. An effective detection method for complex weld defects based on adaptive feature pyramid. In *2023 CAA Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS)*, Yibin, China. Sep 22–24, 2023. IEEE; 2023. p. 1–5. [DOI](#)
2. Li, S. B.; Yang, J.; Wang, Z.; Zhu, S. D.; Yang, G. C. Review of development and application of defect detection technology. *Acta. Autom. Sin.* **2020**, *46*, 2319–36. [DOI](#)
3. Kang, S.; Chen, C.; Zhao, S.; Luo, Y.; Kong, X. Study on infrared image enhancement of wind turbine blades based on adaptive differential multiscale morphology (ADMM). *China. Mech. Eng.* **2021**, *32*, 786–92. [DOI](#)
4. Li, M.; Wang, H.; Wan, Z. Surface defect detection of steel strips based on improved Yolov4. *Comput. Electr. Eng.* **2022**, *102*, 108208. [DOI](#)
5. Ahmed, M. D. F.; Mohanta, J. C.; Sanyal, A. Inspection and identification of transmission line insulator breakdown based on deep learning using aerial images. *Electr. Power. Syst. Res.* **2022**, *211*, 108199. [DOI](#)
6. Yang, Z.; Xu, Z.; Wang, Y. Bidirection-fusion-YOLOv3: an improved method for insulator defect detection using uav image. *IEEE. Trans. Instrum. Meas.* **2022**, *71*, 1–8. [DOI](#)
7. Pan, L.; Chen, L.; Zhu, S.; Tong, W.; Guo, L. Research on small sample data-driven inspection technology of UAV for transmission line insulator defect detection. *Information* **2022**, *13*, 276. [DOI](#)
8. Dian, S.; Zhong, X.; Zhong, Y. Faster R-transformer: an efficient method for insulator detection in complex aerial environments. *Measurement* **2022**, *199*, 111238. [DOI](#)

9. Dong, C.; Zhang, K.; Xie, Z.; et al. Transmission line key components and defects detection based on meta-learning. *IEEE. Trans. Instrum. Meas.* **2024**, *73*, 1–13. [DOI](#)
10. Chang, R.; Xiao, P.; Wan, H.; Li, S.; Zhou, C.; Li, F. A transmission line defect detection method based on YOLOv7 and multi-UAV collaboration platform. *J. Electr. Comput. Eng.* **2023**. [DOI](#)
11. Hao, S.; Ren, K.; Li, J.; Ma, X. Transmission line defect target-detection method based on GR-YOLOv8. *Sensors* **2024**, *24*, 6838. [DOI](#)
12. Wang, Z.; Liu, Z.; Xu, G.; Cheng, S. Object detection in UAV aerial images based on improved YOLOv7-tiny. In *2023 4th International Conference on Computer Vision, Image and Deep Learning (CVIDL)*, Zhuhai, China. May 12–14, 2023. IEEE; 2023. pp. 370–4. [DOI](#)
13. Han, H.; Xue, X.; Li, Q.; et al. Pig-ear detection from the thermal infrared image based on improved YOLOv8n. *Intell. Robot.* **2024**, *4*, 20–38. [DOI](#)
14. Zhao, Z.; Guo, G.; Zhang, L.; Li, Y. A new anti-vibration hammer rust detection algorithm based on improved YOLOv7. *Energy. Rep.* **2023**, *9*, 345–51. [DOI](#)
15. Souza, B. J.; Stefenon, S. F.; Singh, G.; Freire, R. Z. Hybrid-YOLO for classification of insulators defects in transmission lines based on UAV. *Int. J. Electr. Power. Energy. Syst.* **2023**, *148*, 108982. [DOI](#)
16. Dai, Z. Uncertainty-aware accurate insulator fault detection based on an improved YOLOX model. *Energy. Rep.* **2022**, *8*, 12809–21. [DOI](#)
17. Redmon, J.; Farhadia, A. YOLO9000: better, faster, stronger. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, USA. Jul 21–26, 2017. IEEE; 2017. pp. 6517–25. [DOI](#)
18. Redmon, J.; Farhadi, A. YOLOv3: an incremental improvement. *arXiv* **2018**, arXiv:1804.02767. <https://doi.org/10.48550/arXiv.1804.02767>. (accessed 19 May 2025)
19. Tian, Y.; Ye, Q.; Doermann, D. YOLOv12: attention-centric real-time object detectors. *arXiv* **2025**, arXiv:2502.12524. <https://doi.org/10.48550/arXiv.2502.12524>. (accessed 19 May 2025)
20. Liu, S.; Zha, J.; Sun, J.; Li, Z.; Wang, G. EdgeYOLO: an edge-real-time object detector. In *2023 42nd Chinese Control Conference (CCC)*, Tianjin, China. Jul 24–26, 2023. IEEE; 2023. pp. 7507–12. [DOI](#)
21. Wang, C. Y.; Bochkovskiy, A.; Liao, H. Y. M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, Canada. Jun 17–24, 2023. IEEE; 2023. pp. 7464–75. [DOI](#)
22. Ding, X.; Zhang, X.; Ma, N.; et al. RepVGG: making VGG-style ConvNets great again. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, USA. Jun 20–25, 2021. IEEE; 2021. pp. 13728–37. [DOI](#)
23. Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: fully convolutional one-stage object detection. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea. Oct 27 – Nov 02, 2019. IEEE; 2019. pp. 9626–35. [DOI](#)
24. Liu, R.; Lehman, J.; Molino, P.; et al. An intriguing failing of convolutional neural networks and the CoordConv solution. *arXiv* **2018**, arXiv:1807.03247. <https://doi.org/10.48550/arXiv.1807.03247>. (accessed 19 May 2025)
25. Zhang, H.; Zhang, S. Shape-IoU: more accurate metric considering bounding box shape and scale. *arXiv* **2023**, arXiv:2312.17663. <https://doi.org/10.48550/arXiv.2312.17663>. (accessed 19 May 2025)