International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 7850-7861

https://doi.org/10.70517/ijhsa463672

Research on Intelligent Methods for Acceptance of Quality Defects in Distribution Network Line Projects

Bin Feng[1], Keke Lu[2], Shuang Fu[2], Jun Wei[2] and Yu Zou[2,*]
[1] Guangxi Power Grid Co., Ltd., Nanning, Guangxi, 530022, China
[2] Qinzhou Power Supply Bureau of Guangxi Power Grid Co., Ltd, Qinzhou, Guangxi, 535000, China
Corresponding authors: (e-mail: ZY_1106@outlook.com).

Abstract With the continuous development of power system construction and operation, the acceptance of the quality of distribution line engineering has gradually become an important link to ensure the safe and stable operation of the system. In this paper, a defect detection method for distribution line engineering quality based on improved YOLOv5s is proposed. Aiming at the problems of low detection efficiency and large error of the traditional method, the study optimizes the detection performance by introducing a lightweight YOLOv5s model and combining the BiFPN network, the CBAM attention mechanism, and the improved loss function. The experimental results show that the improved model has significant improvement in several performance metrics. The average detection accuracy (mAP) reaches 94.11% when using the model, which is only 0.56% lower than the original YOLOv5s model. In addition, the computational and parametric quantities of the model were reduced by 84.49% and 85.00%, respectively, showing excellent lightweight characteristics. The model is also optimized in terms of detection speed and is able to reach 106.27 frames per second. The conclusion shows that the improved YOLOv5s model performs well in the detection of quality defects in distribution line engineering, which can effectively improve the detection accuracy and shorten the detection time to meet the real-time detection needs of industrial sites.

Index Terms distribution lines, quality defects, YOLOv5s, detection model, lightweight, attention mechanism

I. Introduction

Distribution line engineering refers to the engineering of delivering electricity from transmission lines to the user terminals [1]. In the electric power system, distribution line engineering is a vital link, which bears the important task of transmission lines to deliver electricity to the user's electricity terminal [2], [3]. As a hidden project, the hidden nature of the distribution network line project, the biggest feature of the construction caused by the difficulty, but also on the actual construction of a major test, such as cable landfill acceptance inspection is not possible to dig out for inspection, so the quality of the project is still relying on the construction of the time to ensure more [4]-[7]. At the same time, the current part of the construction unit on the hidden project acceptance work does not pay enough attention to the post-construction inspection without strict implementation of the work requirements, the implementation of the hidden project is no data to the scene to cooperate, the data used by the data clerk is often detached from the actual data, so the hidden project construction naturally does not have enough reliability [8]-[11]. Therefore, the intelligent method of acceptance of quality defects in distribution network line engineering is particularly important [12].

Intelligent acceptance of quality defects in distribution line engineering aims to ensure that the design, construction and operation of the engineering system meet the expected requirements and satisfy the requirements of safety, reliability and high efficiency, which is an important part of ensuring the quality of distribution line engineering [13]-[15]. By following the purpose, principles, methods and processes of the program, it can ensure the safe, reliable and efficient operation of the project [16]. At the same time, it also provides an important reference basis for the subsequent maintenance and management of the distribution network line project.

With the continuous development of the electric power industry, distribution network lines, as an important part of the power system, bear the key tasks of power transmission and distribution. The acceptance of the quality of the distribution line project is crucial for ensuring the safe and stable operation of the power system. The traditional manual detection method has been unable to meet the requirements of modern power engineering for detection efficiency and accuracy.

In recent years, computer vision-based target detection technology has made significant progress in power equipment inspection. The YOLO (You Only Look Once) series of algorithms, as an efficient target detection framework, is widely used in a variety of inspection tasks with its high detection speed and accuracy. In particular,

YOLOv5s, with its lightweight design, is suitable for resource-constrained environments with strong real-time detection capabilities. However, despite its high performance, the YOLOv5s model still faces problems such as large variation in target size and complex background in complex distribution line defect detection tasks. To this end, this paper proposes an improved YOLOv5s model that combines the BiFPN network structure and the CBAM attention mechanism to further enhance the detection accuracy and robustness of the model.The BiFPN network improves the model's detection ability for targets of different scales through multi-scale feature fusion, while the CBAM attention mechanism helps the model to focus on the key regions in the image, which improves the target recognition accuracy. Through these improvements, the model can not only better adapt to the complex detection environment of distribution lines, but also significantly increase the detection speed. Firstly, the diversity and representativeness of the model training data are improved by dataset expansion and labeling. Then, advanced network architecture and data enhancement strategies are used to enhance the model's adaptability to different scenes and complex backgrounds. Finally, the superiority of the improved model in terms of accuracy, efficiency and robustness is verified through a series of experiments, which provides a reliable solution for the intelligent acceptance of quality defects in distribution network line engineering.

II. Improvement of YOLOv5s distribution line engineering quality defect detection model

II. A. YOLOv5s network architecture

II. A. 1) Overall network model for YOLOv5s

YOLOv5 [17] is a target detection framework that has gained a lot of attention in the field of computer vision for its excellent performance and efficient inference speed. In the official code release, there are five different versions of the YOLOv5 model, which are, in descending order of size and complexity of the model, YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. The main difference between these versions lies in the depth and width of the network, i.e., the number of layers in the model and the number of neurons in each layer .

Although these models differ in size, they have the same overall architectural design, but the implementation details of the sub-modules are adapted to suit the needs of models of different sizes. In practice, YOLOv5s is often used for industrial target detection tasks due to its small number of parameters, fast detection speed and high accuracy. Its lightweight design allows YOLOv5s to operate in resource-constrained environments, such as mobile devices or edge computing devices, providing the possibility of real-time detection.

In this study, the YOLOv5s model is used as the basis for its lightweight improvement and optimization for the quality defect acceptance task of distribution line engineering.

II. A. 2) YOLOv5s inputs

YOLOv5 uses a data augmentation method called Mosaic on the input side, which is an extension and improvement of the original CutMix data augmentation method.The core idea of Mosaic data augmentation is to splice multiple images to create more diverse training samples. Specifically, YOLOv5s not only splices four images, but also combines random scaling, random cropping, and random distribution to combine images. This splicing not only increases the diversity of the detection dataset, but also helps to improve the robustness of the model, especially for the detection of small targets. In this way, the model is better able to generalize to targets of different sizes, which in turn improves the detection accuracy.

In the inference phase, YOLOv5s also employs other training strategy tricks. These training strategy techniques enable YOLOv5s to obtain excellent detection results while maintaining high speed inference.

II. A. 3) YOLOv5s backbone network

In the backbone network part of the YOLOv5s model, i.e., the Backbone structure, it inherits the CSPDarknet53 architecture from YOLOv4, an architecture that has maintained a relatively stable design in earlier versions of YOLOv5s. CSPDarknet53 [18] is a deep convolutional network that contains a series of convolutional layers and residual connectivity used to extracting high-level features from the input image.

After v6.0 of YOLOv5s, the new 6×6 convolutional structure can generate feature maps of the same size directly from the original input image without slicing operations. While this change has the same effect as the Focus module in theory, in practice, the new structure may be more efficient for some GPU devices currently on the market. This is an important advantage for application scenarios that require fast response times.

II. A. 4) Neck of YOLOv5s

In the Neck part of the YOLOv5s model, it employs a composite structure with a combination of a feature pyramid network and a path aggregation network. This design aims to fully utilize the feature information of different layers, thus improving the model's ability to detect targets at different scales.

The FPN is a top-down structure that passes the semantic information from the higher layers to the bottom layer, making it possible to obtain rich semantic information for the lower layer feature maps as well. This is particularly important for the detection of small targets, which often have less information in the high-level feature maps. On the contrary, Path Aggregation Network (PAN) is a bottom-up structure that fuses the position information from the bottom layer upwards to help the high-level feature maps recover the spatial details, which is very helpful for the localization of large targets.

Another notable change in the Neck section is that the SPP structure is replaced with the SPPF structure.SPP is a spatial pyramid pooling technique that passes the input through different sized MaxPooling layers in a parallel fashion before fusing the features. This approach can solve the target multiscale problem to some extent as it allows the network to extract information from different receptive field sizes.

YOLOv5s achieves accurate detection of multi-scale targets while maintaining efficient computational performance by combining FPN and PAN as well as employing the SPPF structure in its Neck part of the design. These improvements make YOLOv5s more competitive in the field of target detection, especially when dealing with complex scenes and targets of different sizes.

II. A. 5) Output of YOLOv5s

The Head part of YOLOv5s contains 3 detection layers that receive feature information from the Neck part at 3 different scales, and their outputs are used to detect targets of different sizes to be detected, respectively. Each detection layer is preset with 3 anchors of different aspect ratios, and the target frame regression is computed as follows:

$$\begin{aligned} b_x &= 2\sigma(t_x) - 0.5 + c_x \\ b_y &= 2\sigma(t_y) - 0.5 + c_y \end{aligned} \tag{1}$$

where b_x, b_y are the coordinates of the center point of the prediction box, t_x, t_y are the horizontal and vertical offsets of the center of the prediction target with respect to the upper-left corner of the grid, respectively, c_x, c_y are the coordinates of the upper-left corner of the grid where the prediction center point is located, and σ is the sigmoid activation function.

The predicted target width and height are as in Eq:

$$\begin{aligned} b_w &= p_w * (2\sigma(t_w))^2 \\ b_h &= p_h * (2\sigma(t_h))^2 \end{aligned} \tag{2}$$

b_w, b_h are the width and height of the prediction box, p_w, p_h are the width and height of the anchor box, and t_w, t_h are the coordinates of the center point of the anchor box.

The loss of YOLOv5s consists of three main components, namely classification loss, objectness loss, and localization loss, the first two use BCEloss (binary cross-entropy loss), and the localization loss uses CIoUloss. Classification loss and localization loss are both for positive samples only, and the obj loss is the target bounding box predicted by the network with the true bounding box of the CIoU, for all samples. The loss is calculated as in Eq:

$$Loss = \alpha_1 L_{cls} + \alpha_2 L_{obj} + \alpha_3 L_{loc} \tag{3}$$

where α_1, α_2, and α_3 are equilibrium coefficients.

II. B.Improving the network structure of YOLOv5s

II. B. 1) Improved YOLOv5s

The purpose of introducing the attention mechanism module is to focus on the critical features of the image while suppressing non-critical as well as unnecessary features. The Convolutional Block Attention Mechanism (CBAM) [19], is an attention mechanism module that combines spatial and channel. The CBAM attention module extracts feature information through convolutional operations and mixes cross-channel and spatial information. The module connects a channel attention module to a spatial attention module to allow the neural network to learn the key features from both the channel and spatial dimensions.

The input features of the channel attention module are $F \in \mathbb{R}^{C\times H\times W}$, the one-dimensional convolution $M_c \in \mathbb{R}^{C\times 1\times 1}$ in the channel attention module, the result obtained by convolution is multiplied by the original map,

and the output structure of the channel attention module is used as the spatial attention module's input features, and the 2D convolution module of the spatial attention module performs the convolution operation: $M_s \in \square^{1\times H\times W}$, and finally the output result of the spatial attention module is multiplied with the original map. The structure of the channel attention module is shown in Fig. 1.

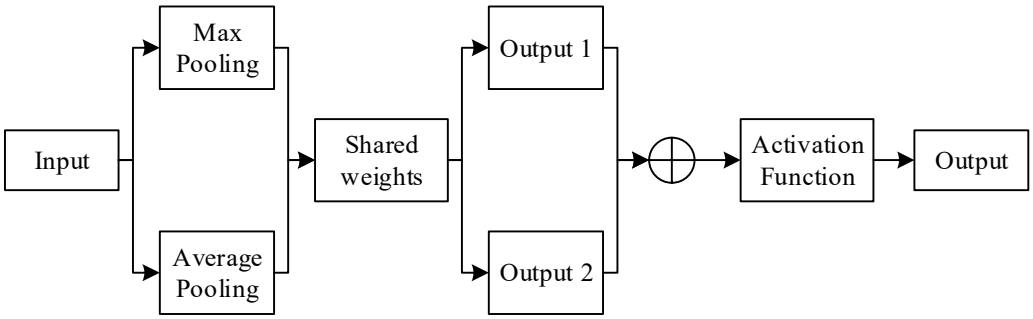

Figure 1: Channel Attention module structure

Among them, the main task of the channel attention module is to extract the useful information of the picture, the input feature map is passed through the parallel maximum pooling layer and the average pooling layer, respectively, to compress the size of the feature map, the feature map is compressed from $C\times H\times W$ to $C\times 1\times 1$, and then the compressed feature map is passed through the shared weighting module, and the channel number of the feature map is The compressed feature map is compressed to $1/r$ of the original size, and then the number of channels is expanded back to the original number of channels C, and 2 results are obtained after the ReLU activation function. The two activation results are added together, and finally a Sigmoid activation function is used to get the final channel attention output, which is multiplied with the input feature map to make the output feature size back to $C\times H\times W$.

The formula for the channel attention module is:

$$M_c(F)=\sigma\left(MLP\left(AvgPool(F)\right)+MLP\left(MaxPool(F)\right)\right) \tag{4}$$

where σ denotes the Sigmoid activation function; MLP denotes the shared weight layer; F denotes the input feature map; $MaxPool$ denotes the channel global maximum pooling; and $AvgPool$ denotes the channel global average pooling.

The structure of the spatial attention module is shown in Figure 2. The spatial attention module is used to better extract the location information of the target, this module will pass the output results of the channel attention through the parallel maximum pooling layer and the average pooling layer, respectively, compress the number of channels of the feature map, C, to 1, and then after the tensor splicing operation will be spliced through the output results of the two pooling layers mentioned above, and then become the feature map of the number of channels, after a convolutional kernel of the size 7×7, into the feature map of the number of channels, then use an S Then a Sigmoid activation function is used to obtain a spatially noted feature map, and finally the output is multiplied with the original feature map to obtain a final output of size $C\times H\times W$.

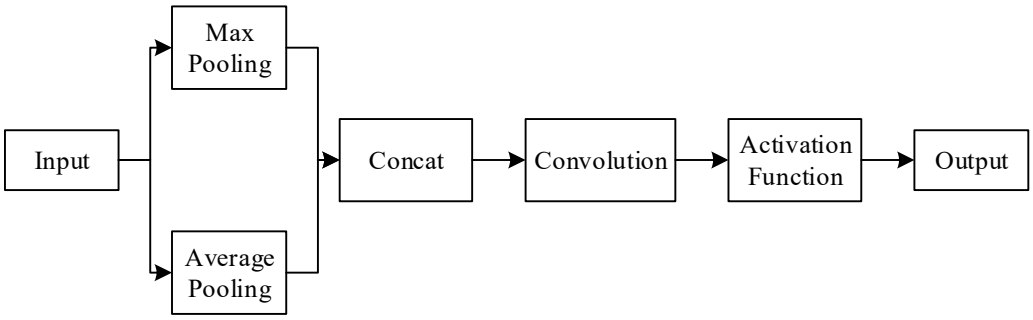

Figure 2: Spatial Attention module structure

The formula for the spatial attention module is:

$$M_s(F) = \sigma\left(f^{7\times7}\left(\left[AvgPool(F); MaxPool(F)\right]\right)\right) \tag{5}$$

where σ denotes the Sigmoid activation function; F denotes the input feature map; $MaxPool$ denotes the spatial global maximum pooling; $AvgPool$ denotes the spatial global average pooling; $f^{7\times7}$ denotes a convolution kernel of size 7×7; and Concat denotes the tensor join operation.

In this study, a C3+CBAM attention focusing module will be added after the CSP. In order to improve the feature extraction ability of the backbone network to the target, the convolutional attention mechanism module of the soft attention mechanism is introduced in the Backbone stage.

II. B. 2) CoT

CoT [20], which can reuse the contextual information between the input keymaps to guide the learning of the dynamic attention matrix, enhances the visual representation. CoT firstly obtains the static contextual representation of the input by convolving the input with the context, and then realizes the dynamic contextual representation of the input by multiplying the obtained attention matrix with the input through 2 consecutive convolutions. Finally, the static and dynamic contextual representations are feature fused for the final output. The CoT structure is shown in Fig. 3.

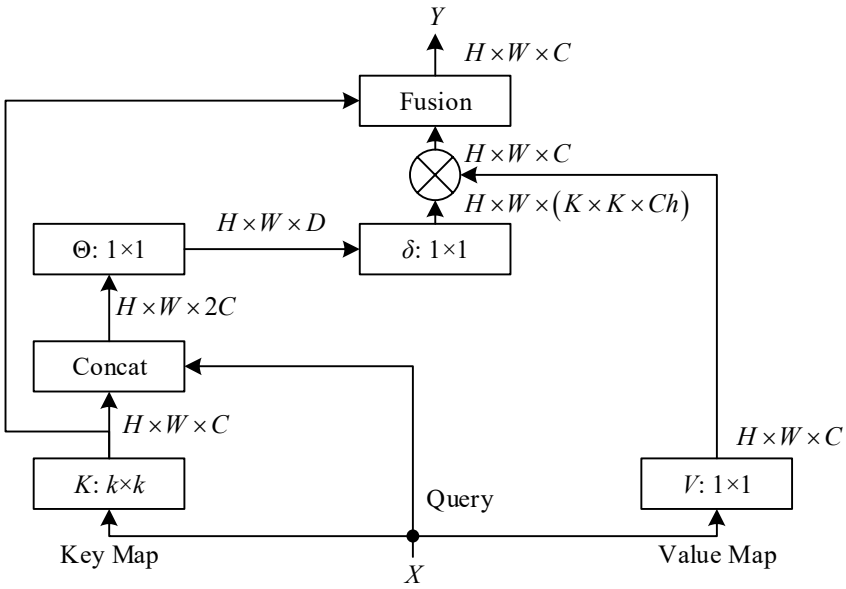

Figure 3: CoT structure

II. B. 3) Improved YOLOv5sNeck

Bidirectional Feature Pyramid Network (BiFPN) introduces learnable weights to learn the importance of different input features, while reusing top-down and bottom-up multiscale feature fusion to act as an efficient multiscale feature fusion.

In order to fuse more feature information without increasing the computational cost, BiFPN will add extra connections between the nodes at the fusion and the input nodes at the same level and with the same resolution of the output feature map. Different input feature maps have different resolutions, and simple splicing cannot distinguish the contribution of different resolution feature maps to the fusion input. Therefore, BiFPN proposes a fast normalized fusion method for different levels of node feature fusion with the following fusion method formula:

$$O = \sum_i \frac{\omega_i}{\varepsilon + \sum_j \omega_j} \cdot I_i \tag{6}$$

where ω_i denotes the feature weights and ε denotes the learning rate. The Relu activation function was used to ensure that each feature weight $\omega_i \geq 0$.

The add connection between the second C3 structure in YOLOv5s Backbone and the 19th layer in Neck improves the feature fusion ability for size targets in the network without changing the parameters of the feature map. The fusion process is formulated as follows:

$$P_{19}^{out} = Conv\left(\frac{\omega_1 \cdot P_6^{in} + \omega_2 \cdot P_{14}^{in} + \omega_3 \cdot P_{18}^{in}}{\omega_1 + \omega_2 + \omega_3 + \varepsilon}\right) \tag{7}$$

where P_i^{out} denotes the output features of the i th layer in the network; P_i^{in} denotes the input features of the i th layer in the network; $Conv$ denotes the convolution operation.

The adaptive spatial feature fusion (ASFF) module weights and fuses the three horizontal feature maps output from PANet respectively, suppresses the inconsistency in the gradient backpropagation process by adding learnable parameters, and makes full use of the features at different scales, so as to achieve the purpose of making more full use of the semantic information of the high-level features in the network and the spatial information in the bottom-level features.

II. B. 4) Improved YOLOv5s Loss Function

The loss-function of the native YOLOv5 is $CIoU$. However, although the $CIoU$ loss adds a measure of the aspect ratio v between the prediction box and the real box on the basis of the $DIoU$ loss, which can accelerate the regression speed of the prediction box to a certain extent, the existing problem is that during the regression process of the prediction box, Once the width-height-aspect ratio of the prediction box and the real box presents a linear proportion, the prediction boxes w and h cannot increase or decrease simultaneously, and thus regression optimization cannot continue. In this paper, the loss-function of YOLOv5 is improved to $EIoU$, which splits the loss term of aspect ratio into the difference between the predicted width and height and the width and height of the minimum external frame respectively, accelerating the convergence of the prediction frame and improving the regression accuracy of the prediction frame. The formula of the $CIoU$ function is as follows:

$$R_{CIoU} = \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \tag{8}$$

where α shows the positive trade-off parameter, which is given by the following equation:

$$\alpha = \frac{v}{(1 - IoU) + v} \tag{9}$$

Here, the formula used to measure the constant prediction frame aspect ratio is as follows:

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{10}$$

It follows that the loss function of CIoU is defined as follows:

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \tag{11}$$

The CIoU Loss though takes into account the overlap area, centroid distance, and aspect ratio of the bounding box regression. But the difference in aspect ratio reflected by v in Eq. (7) is not the real difference between width and height respectively and its confidence level, so it sometimes affects the effective optimization of similarity of the model. To address this problem, the EIou Loss is proposed by splitting the aspect ratios on the basis of CIoU, and its loss function is defined as follows:

$$\begin{aligned} L_{EIou} &= L_{IoU} + L_{dis} + L_{asp} = 1 - IoU \\ &+ \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(b, b^{gt})}{C_w^2} + \frac{\rho^2(b, b^{gt})}{C_h^2} \end{aligned} \tag{12}$$

This loss function contains 3 parts: overlap loss, center distance loss and width and height loss. The first two parts continue the method in CIoU, but the width-height loss directly minimizes the difference between the width and height of the target box and the anchor box, which makes the convergence speed will be faster.

III. Experimental results of the acceptance model for engineering quality defects

III. A. Experimental design and procedure

III. A. 1) Experimental environment and experimental data

The environment configuration for completing the experiments in this paper is Operating System: Window 11, CPU Model: 12th Gen Intel(R) Core(TM) i7-12700H, GPU Model: NVIDIA GeForce RTX 3070 Laptop.The individual networks described in this paper were implemented in Tensor Flow 2.9.1, Cudnn8.2.1, and CUDA11.3 environments.

The experimental data are images taken during the construction and inspection of the distribution network, and the data content is more than 20,000 images of the blocking of low-voltage integrated distribution cabinets in the distribution substation project, and the blocking of cables into and out of cable trenches, tunnels, shafts, buildings, trays (cabinets), and into cable ducts (including used and standby) that conform to the construction quality specifications and the existence of quality defects in the cable laying project.

These data were cleaned by removing damaged images, removing fuzzy images, and removing simple similar images, and 5000 images were selected as the sample database data of cable blocking construction quality in typical distribution network projects, and there were 3000 targets that conformed to the construction quality and 2000 targets that did not conform to the construction quality, and the detection targets in the images were labeled using the labelimg tool. The detection targets in the pictures were labeled using the labelimg tool. In the experiment, all the pictures are divided into training data set and test validation data set in the ratio of 8:2.

III. A. 2) Data set expansion and data set labeling

Since the improved lightweight YOLOv5s used in this paper is trained differently from the improved AODNet network, the training set in this paper is divided into two parts.

This chapter focuses on the study in the improved lightweight network. In order to build a comprehensive and effective training set, 2000 images are selected for image preprocessing in this chapter, and the dataset is expanded to 4500 images by rotating and cropping at different angles, and named as “FogLiftSet” dataset.

Label the target images with LabelImg, an image dataset labeling tool. LabelImg is used to annotate the target images with information such as the category and location of the target frame, and generate the corresponding xml file format. After labeling, according to the YOLO series algorithm dataset format, then according to the ratio of 8:2 between the training set and the test set, the training images are split, and according to the PASCAL VOC dataset training method, the labels, images, and training and test sets are placed in the Annotations, JPEGImages, and ImageSets files respectively, in order to facilitate the training.

III. A. 3) Data Enhancement and Training Strategies

In this paper, the following data enhancement methods are used in the training of YOLOv5s network model: 40% probability of horizontal or vertical flip, 20% probability of panning and small angle rotation for each image during training, and color image enhancement based on Hue-Saturation-Visibility (HSV). The above data enhancement methods can serve to expand the number of training samples, enable the model to learn richer spatial and color information, enhance the model's ability to accurately identify the target objects in the pictures taken in different angles, at different shooting distances, in bright light and dark light, in rain, fog, wind and sand and other adverse weather conditions, and improve the model's robustness in the context of the complex environment.

In this paper, Mixup and Moasic data enhancement methods are comprehensively adopted: the Mixup method is to fuse two pictures together according to a certain degree of transparency; the Moasic method is to stitch one picture with three other random pictures according to a certain picture ratio. The above 2 data enhancement methods enrich the target objects to be detected in different environmental backgrounds, increase the complexity of the data, and further enhance the network's learning of the existing data, thus improving the final detection performance of the model.

In this paper, the following strategies are adopted in the training of the network model: the optimizer adopts the stochastic gradient descent (SGD) algorithm, and the initial training is performed with 4 epochs (training rounds) of warm-up training, after which the learning rate adopts the cosine decay strategy, and the initial values of the hyperparameters, such as the learning rate, are set using the default parameters built into YOLOv5s. AutoAnchor is used to recluster the targets based on the dataset and automatically generate the Anchors template that is more suitable for the dataset in this paper. Multi-scale training is used for training, and 20 epochs of training network model performance indicators are no longer improved as a convergence judgment condition.

The evaluation metrics precision P and recall R are respectively:

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

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

where, T_P is the number of correctly predicted samples in which positive examples are judged as positive examples; F_P is the number of incorrectly predicted samples in which counter-examples are misjudged as positive examples; F_N is the number of incorrectly predicted samples in which positive examples are misjudged as counter-examples.

Average precision (AP) is used as the network performance evaluation index. When the detection contains multiple types of defects, each type of defect generates an AP value, and its summing and taking the average value can get the mean average precision (mAP), mAP is the evaluation of the model's comprehensive defect detection performance in this paper.

$$A_P = \sum_{i=1}^{n-1} (r_{i+1} - r_i) \max(p_{i+1}, p_i) \tag{15}$$

$$m_{AP} = \frac{1}{N} \sum_{1}^{N} A_{Pi} \tag{16}$$

where, $r_i, r_{i+1}, \cdots, r_{n-1}$ are the corresponding Recall values at the interpolated points; p_i is the corresponding Precision value at the interpolated point; N is the number of detected defects species; A_{Pi} is the AP value of the i th defect.

III. B. Experimental results and analysis

III. B. 1) Comparison of Lightweight Models

In this paper, the mainstream lightweight network comparison experiments were conducted, and three convolutional neural network modules, MobileNet V3, GhosNet, and ShuffleNet V2, were selected as the networks for YOLOv5s, and they were compared with YOLOv5s and YOLOv5s+BiFPN in detail, and the results are shown in Table 1.

MobileNet V3, GhosNet and ShuffleNet V2 lightweight networks compared to BiFPN networks in terms of four aspects: average accuracy, number of parameters, computation and model volume, BiFPN networks have the highest average accuracy of detection, and significantly outperform the other network architectures in terms of floating-point operations, computation and model volume, with a number of parameters of only 0.59M,. Demonstrating better lightweight characteristics, this study uses the BiFPN module to optimize the construction of the feature extraction network for YOLOv5s.

Table 1: Comparison of experimental results of light quantitative network

Model	mAP(%)	Parameters/M	FLOPs(G)	size/MB
YOLOv5s	94.61	6.96	15.85	13.73
YOLOv5s+Mobilenet V3	79.33	1.43	2.32	11.11
YOLOv5s+GhosNet	77.40	3.66	7.99	10.06
YOLOv5s+ShuffleNet V2	81.57	1.86	2.23	3.95
YOLOv5s+BiFPN	89.54	0.59	1.87	1.45

III. B. 2) Comparison of Attention Mechanisms

In order to verify the effectiveness of adopting the CBAM module, experimental comparison of the attention mechanisms was carried out, and five mainstream attention mechanisms, CA, ECA, SE, SimAM, and CBAM, were added to the improved YOLOv5s, respectively, with different effects on the four parameters of the model as shown in Table 2.

Adding the CBAM module improves the model performance most significantly, and the average detection accuracy is improved to 88.48%, which is better than the other attention mechanisms. The model size increases by only 1.99 MB, and the number of parameters remains basically unchanged compared to YOLOv5s, YOLOv5s+CA, YOLOv5s+ECA, and YOLOv5s+SimAM.The CA module, although it performs the best in reducing the amount of computation, is not as effective as the CBAM in improving the average detection accuracy. In summary, the introduction of the CBAM module can enhance the model detection capability while maintaining a low computational demand, showing its advantages in building efficient and lightweight target detection models.

Table 2: The attention mechanism compares the experiment

Model	mAP(%)	Parameters/M	FLOPs(G)	Size/MB
YOLOv5s	85.59	0.85	1.82	1.99
YOLOv5s+CA	88.02	0.85	1.80	1.99
YOLOv5s+ECA	85.25	0.86	1.82	2.10
YOLOv5s+SE	86.51	0.91	1.82	2.02
YOLOv5s+SimAM	85.92	0.86	1.89	2.03
YOLOv5s+ CBAM	88.48	0.86	1.84	1.99

III. B. 3) Comparison of ablation experiments

In order to more visually verify the role of each improved module in the YOLOv5s network, the authors added each module in turn for ablation experiments, with the first group being the original YOLOv5s model, and the "√" indicating that the module has been added. The larger the mAP, the better, and the smaller the number of parameters, model size and computation, the better. The comparison results of the ablation experiments are shown in Table 3.

After adding BiFPN network, the model size, the number of parameters and the computational volume are reduced substantially, which greatly reduces the demand for computational resources. After the introduction of CBAM, the average detection accuracy with little change in other metrics is significantly improved by 2.91%. After the introduction of CoT, there is a certain amount of elevation in computation, which increases by 0.73G, but it further improves the average detection accuracy to 90.78%. After adding the ASFF adaptive spatial fusion mechanism in front of the detection head, there is a small increase in the number of parameters and the model volume, but it reduces a certain amount of computation while improving the average detection accuracy. Finally, replacing the EIOU loss function improves the average detection accuracy to 94.1% while other indicators remain unchanged, confirming that the EIOU loss function can improve the detection performance without increasing the model size. After all the modules are replaced and introduced, the model size, number of parameters and computation are only 2.55 MB, 1.11 M and 2.37 G. The above experimental results prove that the improved YOLOv5s network is fully equipped with lightweight and advanced features.

Table 3: The analysis of various modules of ablation experiment

N	BiFPN	CBAM	CoT	ASFF	EIOU	mAP(%)	Parameters/M	FLOPs(G)	Size/MB
1						94.57	7.03	15.80	13.75
2	√					85.60	0.82	1.83	2.02
3	√	√				88.51	0.82	1.80	2.09
4	√	√	√			90.78	0.85	2.53	2.11
5	√	√	√	√		92.56	1.05	2.44	2.59
6	√	√	√	√	√	94.11	1.11	2.37	2.55

III. B. 4) Comparison of Mainstream Algorithms

Fig. 4 shows the PR curve of the original YOLOv5s model and Fig. 5 shows the PR curve of the improved lightweight model. From the comparison of the two PR curves, it can be seen that the overall mAP values for identifying functional defects, structural defects, and external sensory defects are not very much different, and the mAP value of the improved model is 94.11%, while that of the original YOLOv5s model is 94.57%, and the improved model presented in this paper in mAP is only reduced by 0.56%. Therefore, the improved lightweight model can also ensure better detection precision and recall.

In order to further evaluate the target detection capability of the improved YOLOv5s network, the experiments use mAP, number of parameters, computation volume, model volume and detection speed as performance metrics, and compare it with the SSD, Faster-RCNN, YOLOv3-tiny, YOLOv4-tiny, YOLOv5s, YOLOv7-tiny mainstream target detection models, the and YOLOv5s-Sh models are compared, and the detailed results are shown in Table 4.

The YOLOv5s model has a large number of parameters and slow detection speed, and the improved YOLOv5s-Sh model based on ShuffleNet V2 has a 9.13% decrease in average detection accuracy although the number of parameters, computational volume, model volume and detection speed are all reduced significantly, for the improved model in this paper, there is a slight loss in accuracy when compared to the original YOLOv5s network, which is only a decrease of by 0.59% to 94.12%, but significant optimization in the other four key metrics, the amount of parameters is reduced by 85.00% to 1.05M, the amount of computation is reduced by 84.49% to 2.46G, the size of the model is reduced by 81.02%, and the detection speed is increased to 106.27 frames/s. Compared to other

mainstream models, although the average detection accuracy is lower than that of the YOLOv7-tiny, but comprehensively, the improved YOLOv5s algorithm has the smallest number of parameters, computation and model size while maintaining a high average detection accuracy. Through the above comparison, we can clearly see the advantages of this paper's algorithm in lightweight, with good real-time performance, more adaptable to the actual needs of the distribution network line engineering quality defect acceptance scenarios.

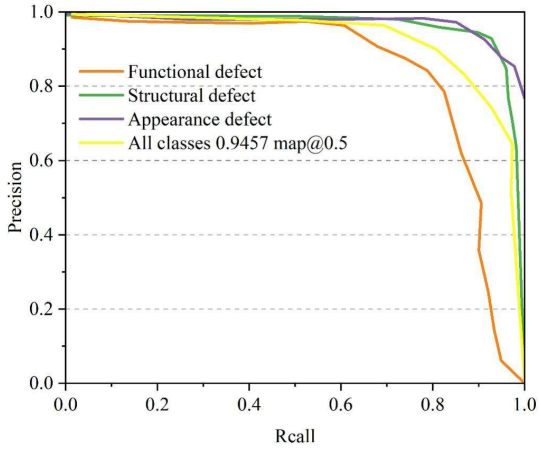

Figure 4: YOLOv5s

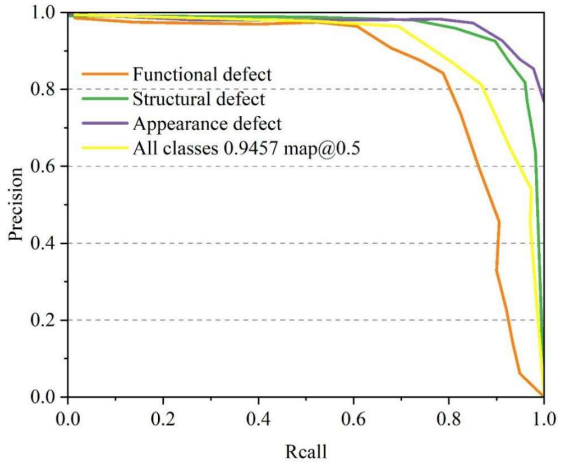

Figure 5: ImprovedYOLOv5s

Table 4: Comparison of experimental results of mainstream algorithm

Algorithm name	mAP(%)	Parameters/M	FLOPs(G)	size/MB	FPS(frame/s)
FasterR-CNN	81.72	137.02	194.3	107.98	65.62
SSD	82.50	23.63	115.68	91.58	83.62
YOLOv3-tiny	88.35	8.59	12.80	16.59	93.66
YOLOv4-tiny	91.88	6.28	15.89	12.90	102.7
YOLOv5s	94.69	7.00	15.86	13.70	75.71
YOLOv7-tiny	96.16	5.98	13.21	11.71	103.56
YOLOv5s-Sh	85.56	0.79	1.87	2.00	108.21
Ours	94.12	1.05	2.46	2.60	106.27

III. B. 5) Model Heat Map Visualization and Analysis

The improved YOLOv5s model has been significantly improved in both detection speed and detection accuracy. After the lightweight improvement, the detection speed of the model is significantly improved, while the accuracy evaluation index mAP gets beyond the expected effect compared with the comparison models such as FasterR-CNN, SSD and YOLOv3-tiny, in this regard, this paper adopts Grad-CAM to carry out the heat map visualization

analysis in order to seek for the explanatory nature, and the heat map visualization analyses of the original model and the improved model are respectively are shown in Fig. 6 and Fig. 7[21]-[23].

The key areas that the network focuses on in the heat map of the network model proposed in this paper are more concentrated, which reflects the network's ability to extract the key information in the picture is stronger, which explains that the method in this paper can improve the accuracy of target detection, and fully reflects the effectiveness of the improved method in this paper.

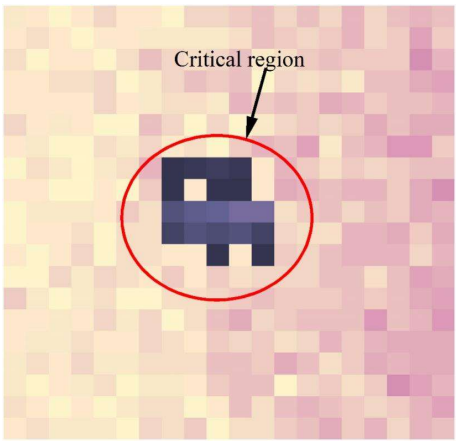

Figure 6: Prototype prediction layer

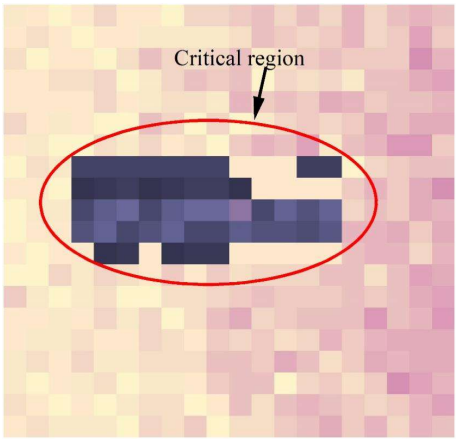

Figure 7: Improved model p4 prediction layer

IV. Conclusion

The improved YOLOv5s model showed significant advantages in the task of detecting quality defects in distribution line engineering. The experimental results show that the detection accuracy (mAP) reaches 94.11% when using this model, which is only a slight decrease (0.56%) compared to the traditional YOLOv5s model. This result indicates that the improved model can further enhance the processing efficiency while ensuring the detection accuracy.

In addition, the improved model is also significantly optimized in terms of computational performance. Compared to YOLOv5s, the amount of parameters of the improved model is reduced by 85.00%, the amount of computation is reduced by 84.49%, and the model size is reduced by 81.02%. These improvements enable the model to run efficiently despite the limited hardware resources and adapt to the actual needs of distribution network line engineering. Most importantly, the detection speed is significantly improved, and a processing speed of 106.27 frames per second can be realized, which fully meets the requirements of real-time detection in industrial sites.

By comparing with other mainstream target detection models, the improved YOLOv5s model performs well in terms of accuracy and efficiency, and its lightweight characteristics make it a strong application prospect. Overall, the intelligent method proposed in this paper provides an efficient, accurate and real-time solution for the acceptance of quality defects in distribution line engineering, which is of great significance for improving the quality management of electric power engineering.

Funding

This research was supported by the Science and Technology Project of Guangxi Power Grid Co., Ltd.: Research on Automatic Acceptance Technology of Distribution Network UAV Based on Visual Fusion Tracking and Image Recognition (040600KC24010001).

References

[1] Pansini, A. J. (2020). Electrical distribution engineering. River Publishers.
[2] Prostejovsky, A. M., Gehrke, O., Kosek, A. M., Strasser, T., & Bindner, H. W. (2016). Distribution line parameter estimation under consideration of measurement tolerances. IEEE Transactions on Industrial Informatics, 12(2), 726-735.
[3] Burke, J. J. (2017). Power distribution engineering: fundamentals and applications. CRC press.
[4] Varela, J., Hatziargyriou, N., Puglisi, L. J., Rossi, M., Abart, A., & Bletterie, B. (2017). The IGREENGrid project: increasing hosting capacity in distribution grids. IEEE Power and Energy Magazine, 15(3), 30-40.
[5] Qu, L., Yu, Z., Song, Q., Yuan, Z., Zhao, B., Yao, D., ... & Zeng, R. (2019). Planning and analysis of the demonstration project of the MVDC distribution network in Zhuhai. Frontiers in Energy, 13, 120-130.
[6] Wang, D., Gao, X., Han, L., Xiao, Z., Feng, T., & Wu, J. (2024, July). Research on Development Direction and Key Technology of Modern Smart Distribution Network. In 2024 7th Asia Conference on Energy and Electrical Engineering (ACEEE) (pp. 195-200). IEEE.
[7] Akimu, N. (2022). Design and Construction of an Automatic Load Monitoring System on a Transformer in Power Distribution Networks. IDOSR Journal of Scientific Research, 7(1), 58-76.
[8] Wang, W., Yu, Y., Xu, J., Zheng, S., & Deng, Z. (2021, March). The application of UAV in intelligent distribution line acceptance system. In 2021 4th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE) (pp. 289-293). IEEE.
[9] Taylor, L., & Short, T. A. (2018). Distribution reliability and power quality. Crc Press.
[10] Razmi, D., Lu, T., Papari, B., Akbari, E., Fathi, G., & Ghadamyari, M. (2023). An overview on power quality issues and control strategies for distribution networks with the presence of distributed generation resources. IEEE access, 11, 10308-10325.
[11] Zhang, M., Bin, M., Song, Y., Liu, W., & Zou, Y. (2025, January). Deep Learning-based UAV Visual Recognition System in Distribution Network Defect Detection. In 2025 International Conference on Electrical Automation and Artificial Intelligence (ICEAAI) (pp. 155-159). IEEE.
[12] Wang, W., Yu, Y., Xu, J., Zheng, S., & Deng, Z. (2021, March). The application of UAV in intelligent distribution line acceptance system. In 2021 4th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE) (pp. 289-293). IEEE.
[13] Yan, Y., Liu, Y., Fang, J., Lu, Y., & Jiang, X. (2021). Application status and development trends for intelligent perception of distribution network. High Voltage, 6(6), 938-954.
[14] Girbau-Llistuella, F., Díaz-González, F., Sumper, A., Gallart-Fernández, R., & Heredero-Peris, D. (2018). Smart grid architecture for rural distribution networks: application to a Spanish Pilot Network. Energies, 11(4), 844.
[15] Saxena, S. N. (2019). Smart Distribution Grid–and How to Reach the Goal. International Journal of Smart Grid, 3(4), 188-200.
[16] Gao, H., Wang, L., Liu, J., & Wei, Z. (2018). Integrated day-ahead scheduling considering active management in future smart distribution system. IEEE Transactions on Power Systems, 33(6), 6049-6061.
[17] Xiaoping Xu,Bingkun Zhou,Wenbo Li & Feng Wang. (2025). A method for detecting persimmon leaf diseases using the lightweight YOLOv5 model. Expert Systems With Applications,284,127567-127567.
[18] Mostafa Farouk Senussi & Hyun Soo Kang. (2024). Occlusion Removal in Light-Field Images Using CSPDarknet53 and Bidirectional Feature Pyramid Network: A Multi-Scale Fusion-Based Approach. Applied Sciences,14(20),9332-9332.
[19] Yusa Chen,Tianhua Meng,Meizhang Wu,Wenya Hu,Dingyi Yang & Wengang Wu. (2025). Nucleobase discrimination based on terahertz spectroscopy using multi-scale convolutional neural network with convolutional block attention module and long short-term memory. Sensors and Actuators: A. Physical,387,116434-116434.
[20] Xuanhong Wang,Cong Li,Zengguo Sun & Luying Hui. (2025). RS-GAN: unsupervised running script font generation via disentangled representation learning and contextual transformer. Pattern Analysis and Applications,28(2),97-97.
[21] H. Wang, J. Wang, L.C. Li, et al. Misalignment tolerance improvement of loosely coupled transformer with ferromagnetic materials via genetic algorithm, Electrical Materials and Applications, 1 (2) e70003, 2024.
[22] Y.T. Wang, X. Li, Z.T. Gao, et al. Analysis and calculation of no-load leakage flux coefficient for arc motors, Electrical Materials and Applications, 2 (1) e70013, 2025.
[23] C.Y. Liu, F.Y. Yang, Y. Han, et al. Advances in high magnetic induction and low loss Fe-based nanocrystalline alloys, Electrical Materials and Applications, 2 (2) e70012, 2025.