



## OPEN ACCESS

EDITED BY  
Hengrui Ma,  
Qinghai University, China

REVIEWED BY  
Linfei Yin,  
Guangxi University, China  
Anton Louise Pernes De Ocampo,  
Batangas State University, Philippines

\*CORRESPONDENCE  
Xu Lu,  
✉ bruda@126.com

RECEIVED 26 August 2023  
ACCEPTED 23 October 2023  
PUBLISHED 03 November 2023

CITATION  
Liu J, Hu M, Dong J and Lu X (2023), The application of a lightweight model FA-YOLOv5 with fused attention mechanism in insulator defect detection. *Front. Energy Res.* 11:1283394. doi: 10.3389/fenrg.2023.1283394

COPYRIGHT  
© 2023 Liu, Hu, Dong and Lu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# The application of a lightweight model FA-YOLOv5 with fused attention mechanism in insulator defect detection

Jun Liu<sup>1,2</sup>, Mingming Hu<sup>1</sup>, Junyuan Dong<sup>1</sup> and Xu Lu<sup>1,3\*</sup>

<sup>1</sup>Institute of Automation, Guangdong Polytechnic Normal University, Guangzhou, China, <sup>2</sup>Guangzhou Key Laboratory of Intelligent Building Equipment Information Integration and Control, Guangzhou, China, <sup>3</sup>Pazhou Lab, Guangzhou, China

Insulators are important components of transmission lines, serving as support for conductors and preventing current backflow. However, insulators exposed to natural environments for a long time are prone to failure and can cause huge economic losses. This article proposes a fast and accurate lightweight Fast and Accurate YOLOv5s (FA-YOLO) model based on YOLOv5s model. Firstly, attention mechanisms are integrated into the network module, improving the model's ability to extract and fuse target features. Secondly, the backbone part of the network is lightweightened to reduce the number of parameters and computations at the cost of slightly reducing the accuracy of detecting a few objects. Finally, the loss function of the model is improved to accelerate the convergence of the network and improve detection accuracy. At the same time, a visual insulator detection interface is designed using PyQt5. The experimental results show that the algorithm in this paper reduces the number of parameters by 28.6%, the computational effort by 35.7%, and the mAP value by 1.7% compared with the original algorithm, and is able to identify defective insulators quickly and accurately in complex backgrounds.

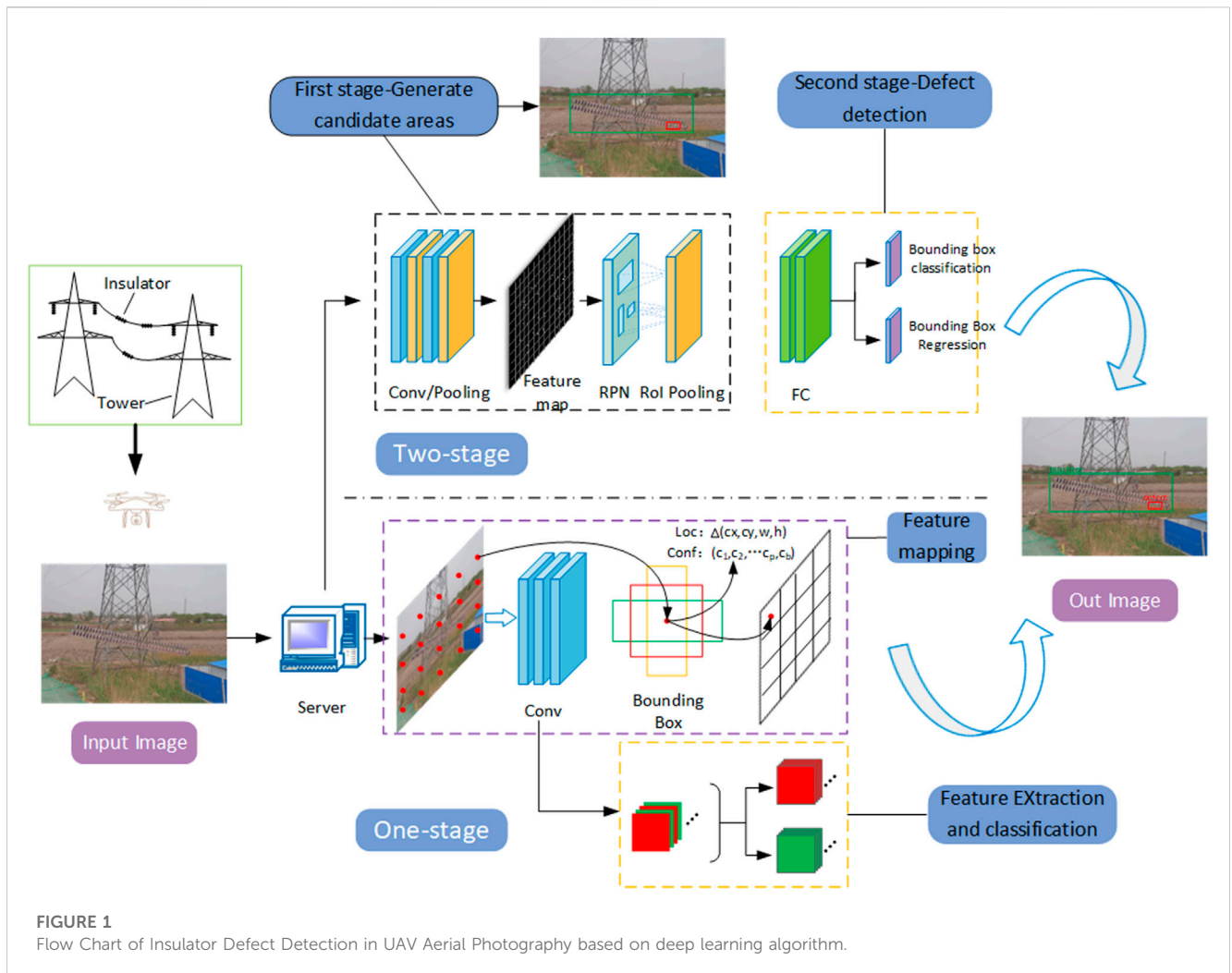
## KEYWORDS

insulators, defect detection, attention mechanism, lightweighting, WIoU\_Loss

## 1 Introduction

With the increasing demand for electricity, transmission lines have spread all over the country. Insulators, as an important part of them, have good mechanical support and electrical insulation properties, and play an important role in supporting the conductor and preventing the current from returning to the ground during the whole transmission process. However, its long-term exposure to strong electric field environment and susceptibility to adverse weather conditions such as rain, snow, and extreme temperatures, resulting in defects such as spontaneous explosion and fracture, creates a huge potential risk to the safe and stable operation of transmission lines, and according to statistics, the highest number of failures in power systems is caused by insulator defects (Chen, 2020; El-Hag, 2021). Therefore, fast and accurate detection of defective insulators and timely replacement are particularly important for the safe operation of the entire transmission system.

The defect detection of insulators is mainly divided into insulator localization as well as defect detection. The defect detection of insulators can be divided into manual observation, traditional image processing based and deep learning based methods. Among them, the manual observation method is time-consuming and labor-intensive, and has certain safety



risks (Yu et al., 2019). And the traditional image processing-based methods need to set the target features artificially, and different features need to be set for different targets, and the recognition accuracy is low, which can easily cause false detection or missed detection. In (Lu et al., 2017), an active contour model is proposed for insulator segmentation based on the shape and texture features of insulators, and the method proves to be effective in identifying defective insulators even in a cluttered background. Zhang et al. (2018) proposed a computer vision-based insulator feature extraction method, which extracts texture features through a grayscale co-occurrence matrix and then detects insulator features using local features. Although the traditional image processing-based method has been able to detect defective insulators well, the detection process is complicated and easily disturbed by the background environment, resulting in missed and false detection. To overcome the interference of complex background, Zheng H. et al. (2020) proposed an improved infrared insulator image detection model based on the complex substation environment, which improves the extraction capability for insulator infrared image features by generating new feature pyramids with feature enhancement modules. However, the infrared imaging-based method is susceptible to the influence of temperature leading to poor detection results.

The above methods based on traditional image processing can only accurately identify defective insulators in a specific environment because they cannot automatically extract insulator features, but insulators are usually in complex background environments such as rivers, farmlands, construction sites and forests, so a method that can automatically extract insulator feature information from images is urgently needed.

With the rise of deep learning technology, target detection algorithms have achieved great success in the field of insulator defect detection by virtue of their fast and accurate recognition capability. Compared with traditional image processing methods, deep learning-based target detection algorithms can automatically extract deep feature information in images, reduce recognition time and improve detection accuracy (Yang et al., 2021). The flow of insulator defect detection based on deep learning algorithm is shown in Figure 1.

At present, the mainstream object detection algorithms are mainly divided into Two stage and One stage. Two stage first generates a prior box based on the target object, and then recognizes and judges the objects within the prior box. This method has high detection accuracy and can accurately identify the target, but the detection time is long. The mainstream algorithms include Faster R-CNN (Ren et al., 2015) and Mask R-CNN (He et al.,

TABLE 1 Relevant datasets used to cite the Reference.

Author	Number of data	Method	Data enhancement	Fault location
Shuang et al. (2023)	806	Detail R-CNN	×	√
Zhao et al. (2021)	10,468	Faster R-CNN	×	√
Yi et al. (2023)	1700	YOLOv5s	√	√
Zhang et al. (2023)	848	YOLOv4	√	√
Miao et al. (2019)	6700	SSD	×	√
Chen Y et al. (2023)	2448	YOLOv8n	√	√

2017). For example, Shuang et al. (2023) introduced a feature enhancement and assisted classification module based on Faster R-CNN to improve the accuracy of model detection. The data enhancement method of YOLOv5-X was also ported to expand the dataset. Zhao et al. (2021) firstly used feature pyramid network to improve the Faster R-CNN model, and then segmented the image by hue, saturation and value color space (HSV) adaptive thresholding algorithm, and finally localized and detected the defective insulators. Tan et al. (2022) used Mask R-CNN model to segment out insulators and detected defects such as breakage, dirt, foreign matter and flashover by multi-feature fusion and cluster analysis model.

One stage directly locates and recognizes targets, which has a fast detection speed and can achieve real-time detection. However, the detection effect is not satisfactory. The mainstream algorithms include YOLO series algorithms and SSD (Wei et al., 2016) algorithm, and timely detection and replacement of defective insulators is important for the safe and stable operation of transmission lines. In order to detect the working status of insulators in real time, Yi et al. (2023) improved the Neck part of the YOLOv5s model and proposed a new attention module MainECA to enhance target perception, and the proposed YOLO-Small model reduced the number of parameters while improving the detection accuracy. Zhang et al. (2023) used GhostNet as the Backbone network of the YOLOv4 model, and at the same time optimized the model using K-means algorithm and Focal loss function. Chen Y et al. (2023) added the GSConv module to the latest YOLOv8n algorithm to reduce the complexity of the network, and also adopted a lightweight Content-Aware Feature Reconstruction (CARAFE) structure to enhance the feature fusion capability of the model. Miao et al. (2019) used a combination of SSD model and two-stage fine-tuning strategy to complete the detection of defective insulators, which can automatically extract multi-level features of images and can identify porcelain insulators and composite insulators quickly and accurately in complex backgrounds.

Based on the fact that deep learning methods need to use a large number of datasets to achieve better results, and then there are not many open-source insulator datasets due to confidentiality factors, most of the methods mentioned above use data augmentation strategies to expand their datasets, as shown in Table 1.

In Table 1, Shuang et al. used 806 images captured from Guangxi Power Grid in China as the dataset, and did not expand the dataset using image processing methods, but directly used the data enhancement methods in YOLOv5x to enhance the training data. Zhao et al. used 4 datasets with a total of 10,468 images, and Yi et al.

used data enhancement methods such as rotating, panning, scaling, cropping, etc. to expand 1700 original images to 5180 images. Zhang et al. also used the above methods to expand 848 images in the original open-source Chinese Power Line Insulator Dataset (CPLID) into 880 images. Zhang et al. also used the above method to expand 848 images in the original open-source Chinese Power Line Insulator Dataset (CPLID) to 5832 images, and Miao et al. used a drone to take 6700 original images on the transmission line as a dataset for their experiments. Finally, Chen et al. expanded the open-source datasets CPLID and Insulator Defect Image Dataset (IDID) to 5676 images using common data expansion methods.

The main research objective of this article is to propose an improved YOLOv5 algorithm, FA-YOLOv5s, to address the issues of high computational complexity, slow detection speed, complex background, mutual occlusion, and small targets in the existing insulator defect detection algorithms. The proposed method mainly improves the network structure of YOLOv5 model and loss function, so that the new algorithm can quickly and accurately identify insulators in complex environments and detect whether they have faults. The main contributions of this article are as follows: 1) Integrating the Convolution Block Attention Module (CBAM) (Woo et al., 2018) with the network's C3 module enhances the network's ability to fuse insulation feature information, improving detection accuracy. 2) By using Partial Convolution (PConv) to lightweight the main network part of the model, the computation cost is reduced at the cost of reduced accuracy. 3) The loss function of the network was improved by using Wise\_IoU Loss as the loss function, which improved the convergence speed of the model.

## 2 Related work

As one of the current popular target detection methods, the YOLOv5 algorithm is a product of continuous innovation and improvement based on the YOLOv3 (Redmon and Farhadi, 2018) and YOLOv4 (Bochkovskiy et al., 2020) algorithms. It combines the advantages of both algorithms, has fewer parameter quantities, and a simpler structure. While accelerating the detection speed, it also increases the detection accuracy, and achieves better detection results on PASCAL VOC (Everingham et al., 2015) and COCO (Lin et al., 2014) datasets. According to its network depth and width, YOLOv5 is successively YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, with the fastest detection speed and lowest accuracy. The comparison on the COCO datasets is shown in Table 2.

TABLE 2 Comparison of YOLOv5 parameters in COCO dataset.

Network model	Image size	mAP 0.5	mAP 0.5:0.95	Speed (v100/ms)	Parameter(M)
YOLOv5s	640	55.4	36.7	2.0	7.3
YOLOv5m	640	63.1	44.5	2.7	21.4
YOLOv5l	640	66.9	48.2	3.8	47.0
YOLOv5x	640	68.8	50.4	6.1	87.7

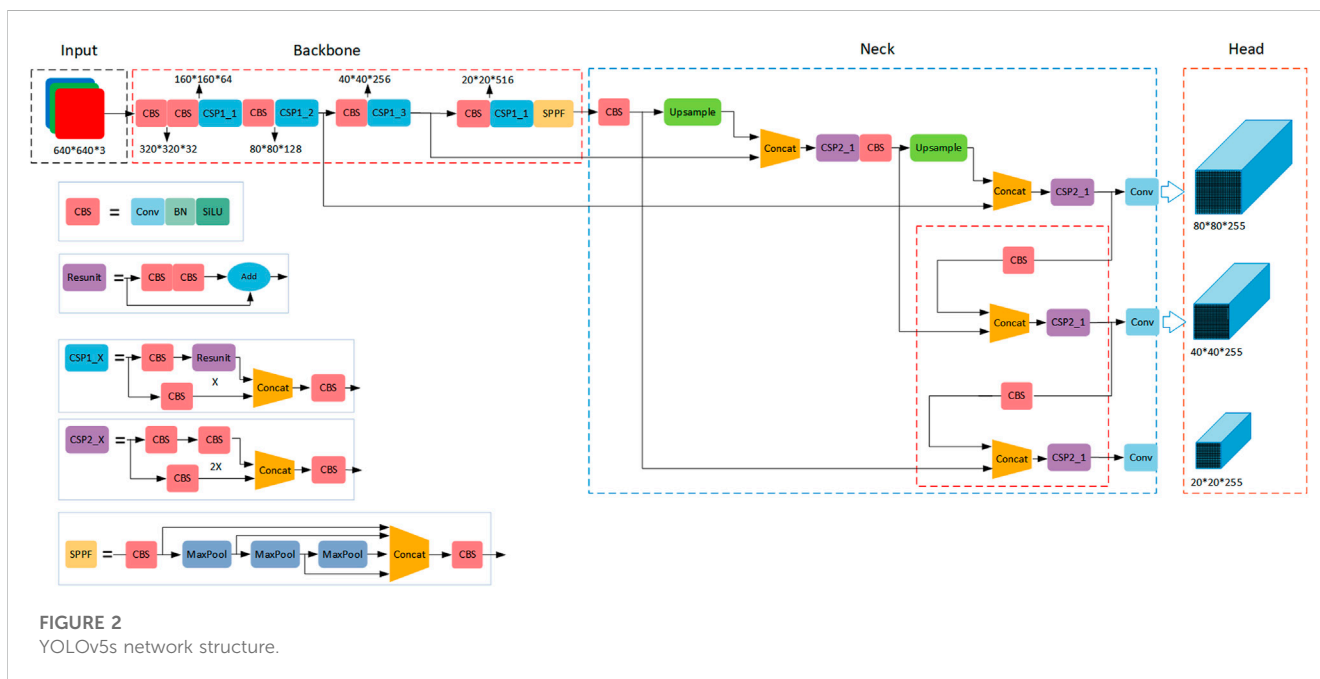


FIGURE 2 YOLOv5s network structure.

The network model is shown in Figure 2. YOLOv5s network structure is mainly divided into four parts: Input, Backbone, Neck and Head. Mosaic (Lewy and Mańdziuk, 2023) data enhancement is performed on the input side to speed up the image processing and reduce the memory size of the model, which makes the model obtain better detection results. backbone is mainly composed of Conv-Batch Normalization-SiLU(CBS), Stack 3 convolutional layers on top of multiple bottleneck layers (C3) and Spatial Pyramid Pooling Fusion (SPPF) modules. The CBS module consists of a normal convolutional layer (Convolutional), a batch normalization (Batch Normalization) and an activation function (SiLU), while C3 consists of three standard convolutional layers and several Bottleneck modules, which are structured The SPPF is divided into two branches, one passing through multiple Bottleneck stacks and 3 standard convolutional layers, and the other passing through a basic convolutional module, and finally they are concatenated. SPPF improves the perceptual field of the network through feature extraction with maximum pooling of different pooling kernel sizes. The main role of the Neck part is to deep fuse the features extracted from the Backbone The Head part outputs the input size of  $640 \times 640$  images as  $20 \times 20$ ,  $40 \times 40$  and  $80 \times 80$  size feature maps, which are used to predict large, medium and small targets in three different sizes.

### 3 Fast and accurate FA-YOLOv5 algorithm

Current improvements to the YOLOv5 algorithm focus on improving the accuracy and convergence speed, while ignoring the complexity of the network model and the increase in the number of parameters, e.g., Han et al. (2022) added the ECA-Net attention mechanism to the backbone feature extraction network of YOLOv5, and also used a bidirectional feature fusion network in the feature fusion layer to enhance the detection of small targets. Gao et al. (2021) proposed a convolutional attention module with batch normalization (BN-CBAM) and a multi-level feature fusion module to enhance the detection of small targets. Although these methods are effective in improving the detection accuracy for small targets, they also make the network structure more complex and reduce the detection speed. In this paper, the convolutional attention mechanism CBAM module is fused with the C3 module of Neck part to improve the accuracy of detection. At the same time, the network is lightweighted to address the problems of complex network structure, number of parameters, and large computation. Finally, the latest WIoU loss is used as the loss function of the model, which speeds up the convergence, makes full use of the dynamic non-monotonic FM potential, and solves the problem of unbalanced sample quality.

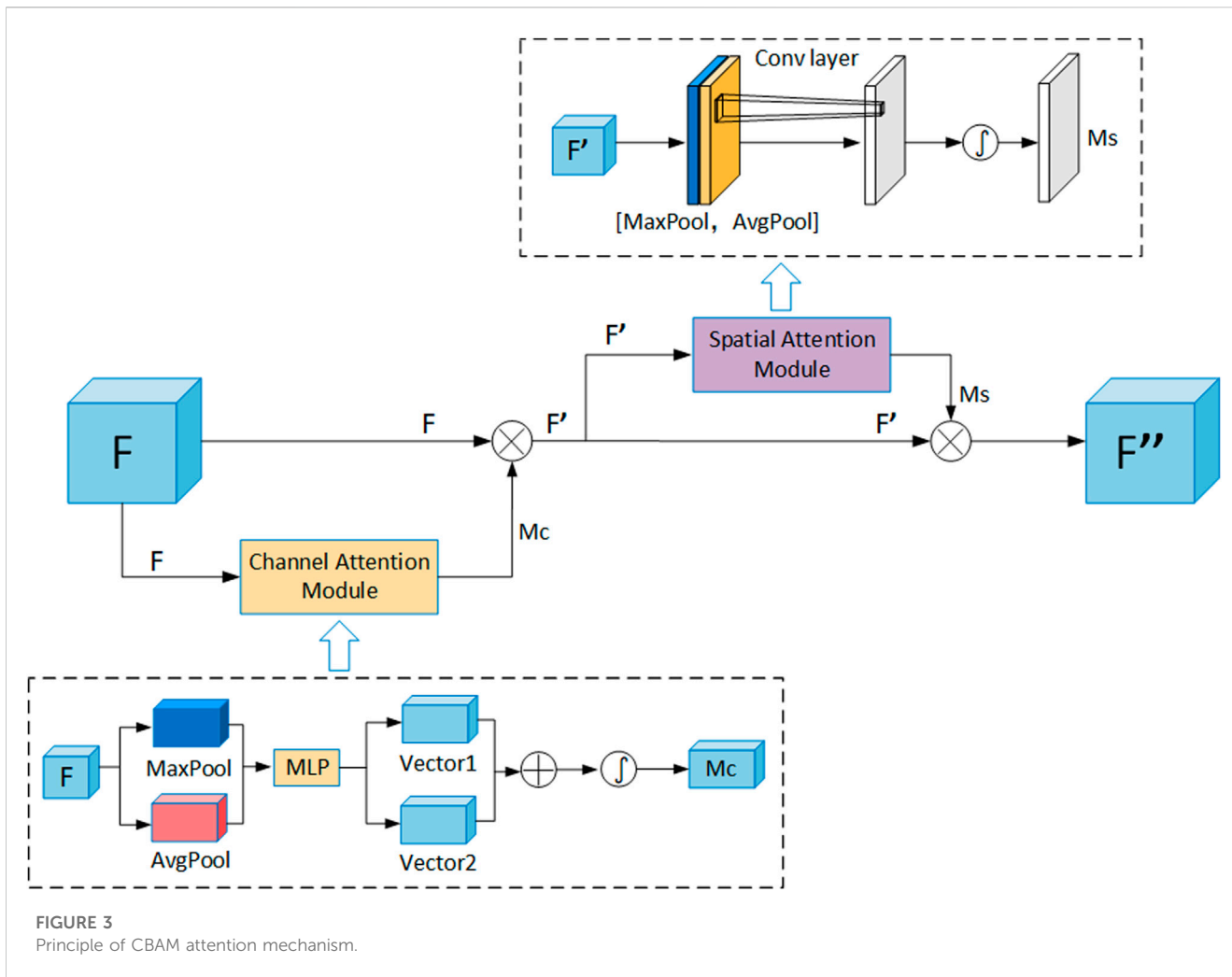


FIGURE 3 Principle of CBAM attention mechanism.

### 3.1 Incorporating attention mechanism in C3 module

Since insulators are mostly in complex backgrounds and the defective part of insulators is a relatively small part of the whole image, it is difficult for the algorithm to extract feature information of insulators and their defects effectively. In order to enhance the extraction of target feature information, researchers proposed the attention mechanism (Vaswani et al., 2017), whose main role is to enhance the extraction of various appearance features of the target and make the algorithm biased to extract the features, the core of which is to make the network focus on the region of the target in the image rather than the whole image. By making the algorithm focus on the feature information of the target and ignore other unimportant information to improve the detection performance of the algorithm, the attention mechanism has been widely used in computer vision tasks such as target detection and image segmentation in recent years, and occupies an important position in the field of deep learning.

Attention mechanisms are usually divided into channel attention mechanisms and spatial attention mechanisms, which focus on the channel dimension and spatial dimension, respectively. Channel attention is used to deal with the

assignment relationship of feature map channels, while spatial attention allows neural networks to focus more on target regions in the image and ignore irrelevant regions, and simultaneous attention allocation to both dimensions enhances the effect of attention mechanisms on model performance.

The workflow of CBAM is shown in Figure 3, where the feature map is first passed through the channel attention module, then the feature map is multiplied with the channel weights and input to the spatial attention module, and finally the normalized spatial weights are multiplied with the feature map input to the spatial attention module to obtain the final weighted feature map. The final weighted feature map is obtained. This module not only saves parameters and computational effort, but also can be easily added to other network structures. For example, Wang et al. (2022) directly added the CBAM attention module to the YOLOv5s network structure to improve the insulator feature extraction capability and achieve insulator detection in complex backgrounds, but the method is not effective for insulator defect detection of small targets.

The overall formula of Figure 3 is shown in Eqs 1, 2.

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \quad (1)$$

$$M_s(F) = \sigma(f^{\text{conv}}([\text{AvgPool}(F); \text{MaxPool}(F)])) \quad (2)$$



TABLE 3 Comparison of different attention mechanisms and addition methods.

Methods	Precision (%)	Recall (%)	mAP0.5 (%)	Parameters(M)	GFLOPs
YOLOv5	95.5	94.5	95.4	7.0	16.0
NAM	96.1	94.0	95.2	7.0	16.0
CBAM	96.5	94.7	95.9	7.0	16.0
ECA	95.4	95.3	95.5	7.0	16.0
CoordAtt	<b>96.8</b>	95.7	96.3	7.0	16.0
GAM	96.7	95.1	96.4	7.5	16.3
C3CBAM	96.7	<b>96.1</b>	<b>96.5</b>	<b>6.9</b>	<b>15.7</b>

The bold portion of the table indicates the value with the best performance in the metric.

$M_c(F)$  and  $M_s(F)$  are the channel and spatial attention weights of the feature layer  $F$ , respectively; AvgPool and max Pool are the average pooling and maximum pooling operations; MLP stands for multilayer perceptron;  $\sigma$  is the sigmoid activation function. as the input to the next layer, as shown in Eq. 3.

$$\begin{cases} F' = M_c(F) \otimes F \\ F'' = M_s(F') \otimes F' \end{cases} \quad (3)$$

where  $\otimes$  denotes element-wise multiplication,  $F$  is the intermediate quantity of the feature layer passing through the channel attention module. and  $F''$  denotes the output passing through the spatial attention module.

The current common method for improving attention mechanisms is to directly add them to the network structure, which does not fully leverage the effectiveness of attention mechanisms. Although this approach does improve the detection accuracy to some extent, it also increases the number of layers and parameters in the network. To further reduce the number of model parameters and fully leverage the effectiveness of attention mechanisms, this article combines attention mechanisms with the C3 module to form a new module, namely, C3CBAM. At the same time, the newly generated C3CBAM module further enhances the model's capability to focus on target feature information. This module strengthens the model's ability to fuse and extract target feature information from both channel and spatial dimensions, allowing for accurate identification of target feature information even in complex background environments. As a result, efficient insulation defect detection can be achieved.

To verify the effectiveness of this method, we conducted experiments by adding different attention mechanisms after the same C3 layer and compared them with the method of incorporating CBAM into the network layer. The experimental results are shown in Table 3. We added different attention mechanisms to the network model for comparison experiments, which were conducted after adding different attention mechanisms to the same Conv layer while ensuring that the number of other parameters of the experiment was the same. As can be seen from Table 3, different methods have different effects on the performance of the original model, Normalization-based Attention Module (NAM) and Efficient Channel Attention (ECA) reduce the accuracy of the model detection. The other attention mechanisms all have some improvement effect on the detection performance of the model, among which the Global Attention Mechanism (GAM) attention

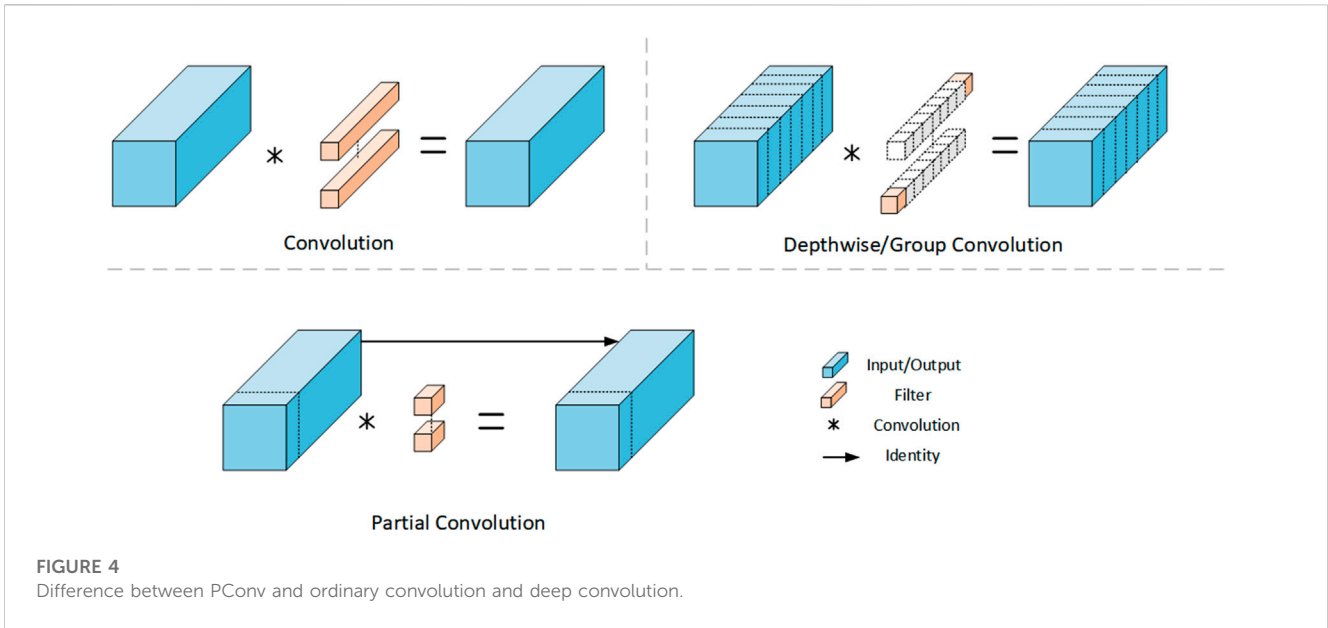
mechanism has the biggest improvement effect, but it increases the number of parameters and computation of the model, because the purpose of this study is for fast and accurate insulator defect detection algorithm, out of the comprehensive considerations, we choose to integrate the CBAM Attention Mechanism and C3 module fusion method to improve the original model. This method reduces the number of parameters and computation to some extent, and most importantly has the highest performance enhancement effect on the original model.

### 3.2 Lightweight network architecture

The FasterNet (Chen J et al., 2023) network recently released by CVPR far exceeds other existing networks in terms of lightweight as well as the balance of detection performance. The current mainstream lightweight networks such as MobileNet, ShuffleNet, and GhostNet utilize deep convolution (DWConv) or group convolution (GConv) to extract spatial feature information, which although greatly reduces the number of parameters and floating point operations (FLOPs), but the computation is not efficient, increases the number of layers of the network, runs slower, and greatly reduces the accuracy and effectiveness of detection, while adding some additional data operations. In order to maintain high accuracy while reducing FLOPs, Chen et al. proposed local convolution (PConv), which works as shown in Figure 4.

PConv has lower computational effort as well as higher computational efficiency, which can utilize the computational power of the device more efficiently and also improves the model's ability to extract spatial feature information. Based on this, Chen Y et al. (2023) proposed FasterNet, which can achieve better results in classification, detection and segmentation tasks at a faster rate, and its can replace the Backbone part of the YOLOv5 model.

This article improves the backbone network of YOLOv5 using PConv, FasterNet, MobileNet, ShuffleNet, and GhostNet respectively. Through experimental comparisons, it is shown that PConv effectively reduces the complexity and parameter count of the network while maintaining high accuracy. The experimental results are shown in Table 4. The backbone network using the FasterNet improvement algorithm has the highest detection accuracy but the number of parameters is still high, while replacing the entire backbone part of the network using ShuffleNetV2 greatly reduces the number of parameters and computation, but also increases the number of layers of the network, and the detection speed is also reduced.



**FIGURE 4** Difference between PConv and ordinary convolution and deep convolution.

**TABLE 4** Experimental results of different ways of light weight treatment.

Methods	Layers	Parameters(M)	GFLOPs	Precision (%)	mAP0.5 (%)	FPS
PConv	<b>129</b>	5.1	10.5	95.1	94.4	<b>89</b>
Faster-Net	228	6.4	14.0	95.4	<b>95.2</b>	76
MobileNetV3	320	1.4	2.3	94.8	93.2	80
ShuffleNetV2	193	<b>0.8</b>	<b>1.9</b>	94.5	92.9	82
GhostNet	500	5.3	8.4	<b>96.8</b>	94.1	73

The bold portion of the table indicates the value with the best performance in the metric.

MobileNetV3 and GhostNet both reduce the number of network parameters and computation at the cost of increasing the number of network layers. Based on this, this paper uses PConv to lighten the Backbone of the network, which can greatly reduce the number of parameters and computation of the model at the cost of a small reduction in accuracy.

### 3.3 Improvement of the loss function

The loss functions of YOLOv5 model are Classification loss, Localization loss and Confidence loss, and the sum of the three loss functions is the size of the total loss function. The calculation formula is as in Eq. 4.

$$\text{Loss} = \text{loss}_{\text{box}} + \text{loss}_{\text{obj}} + \text{loss}_{\text{cls}} \quad (4)$$

IoU\_Loss (Yu et al., 2016) is the first proposed loss function for target detection, but it only considers the overlap area of the detection frame and the target frame, which has certain defects. the appearance of GIoU\_Loss (Rezatofighi et al., 2019) loss function solves the shortcomings of IoU\_Loss to a certain extent, but it also has the disadvantages of not accurate enough boundary regression and slow convergence speed. The subsequent DIoU\_Loss (Zheng Z.

et al., 2020) loss function takes the overlap area and centroid distance into account and accelerates the convergence speed, but does not take the aspect ratio factor into account. To address these drawbacks, the CIoU\_Loss (Zheng et al., 2021) loss function takes into account the overlap area, centroid distance and aspect ratio influence factor  $\alpha$  and  $v$ , and its calculation process is shown in Eq. 5.

$$\text{CIoU} = 1 - \text{IoU} + \frac{\rho^2(\mathbf{b}, \mathbf{b}^{\text{gt}})}{C^2} + \alpha v \quad (5)$$

where,  $\mathbf{b}$  represents the center coordinates of the prediction frame,  $\mathbf{b}^{\text{gt}}$  represents the parameter of the center of the real target bounding box.  $\rho^2$  represents the Euclidean distance between the two centroids,  $c$  represents the minimum external rectangle diagonal length of the two rectangles,  $v$  is used to measure the consistency of the aspect ratio, and  $\alpha$  is the weight function. The values of  $\alpha$  and  $v$  are shown in Eq. 6.

$$\begin{cases} \alpha = \frac{v}{(1 - \text{IoU}) + v} \\ v = \frac{4}{\pi^2} \left( \arctan \frac{m^{\text{gt}}}{n^{\text{gt}}} - \arctan \frac{m}{n} \right)^2 \end{cases} \quad (6)$$

The original YOLOv5 algorithm uses CIoU\_Loss as the loss function of the network, however, the  $v$ -value used to measure the

aspect ratio is too complex and slows down the convergence to some extent, so when one of the two variables increases (shrinks), the other one will shrink (increase). To solve this problem, (Zhang et al., 2022), proposed EIoU Loss by splitting the aspect ratio on the basis of CIoU, which accelerated the speed of convergence and improved the accuracy of regression. Focal-EIoU was also proposed to focus on high-quality anchor frames, which optimized the problem of sample quality imbalance in the regression task and made the regression process more focused on high-quality anchor frames, and the calculation process of EIoU\_Loss is shown in Eq. 7. Thus (Yang et al., 2022) used EIoU as a loss function to improve the YOLOv3 algorithm, which improved the overlap between the predicted and actual frames of the target and accelerated the convergence speed.

$$\begin{aligned} L_{EIoU} &= L_{IoU} + L_{dis} + L_{asp} \\ &= 1 - IoU + \frac{\rho^2(b, b^{gt})}{C^2} + \frac{\rho^2(w, w^{gt})}{C_w^2} + \frac{\rho^2(h, h^{gt})}{C_h^2} \end{aligned} \quad (7)$$

Although Focal-EIoU solves the problem of sample quality imbalance to some extent, the potential of non-monotonic FM is not fully utilized due to its static focusing mechanism (FM), so (Tong et al., 2023) proposed an IoU-based loss with dynamic non-monotonic FM, namely, Wise IoU (WIoU), which has a bounding box regression of attention-based loss WIoU v1, WIoU v2 with non-monotonic FM, and WIoU v3 with dynamic non-monotonic FM. In this paper, WIoU v3 is used as the loss function of the network, and its gradient gain allocation strategy with dynamic non-monotonic FM is utilized to trade-off the learning ability of high quality as well as low quality samples and improve the overall performance of the model. The calculation formula is shown in Eq. 8.

$$\begin{aligned} L_{WIoUv1} &= R_{WIoU} L_{IoU} \\ R_{WIoU} &= \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 - H_g^2)^*}\right) \end{aligned} \quad (8)$$

where  $W_g$ ,  $H_g$  denote the width and height of the minimum enclosing frame. To prevent  $R_{WIoU}$  from creating gradients that hinder convergence,  $W_g$  and  $H_g$  are separated from the computational graph (the superscript \* indicates this operation). No new metric like aspect ratio is introduced because it effectively eliminates the factors that hinder convergence. To significantly amplify the localization loss (LIoU) of the normal quality anchor box, the range of  $R_{WIoU}$  is [1,e] while the range of LIoU is [0,1], which will significantly reduce the Rwiou of the high quality anchor box and focus on their centroid distance when the anchor box overlaps with the target box.

### 3.4 Network structure of this paper

In order to reduce the complexity of the model and make it more suitable for deployment on mobile devices such as UAVs, this paper uses PConv to lighten the backbone part of the network. At the same time, CBAM attention is fused with C3 module to give full play to CBAM's ability to extract target feature information in channel and space, which improves the accuracy of detection. Finally, WIoU\_loss is used as the loss function of the network, and the improved part is shown in red, and the specific network structure is shown in Figure 5.

- (1) Backbone: Compared with the old version of Spatial Pyramid Pooling Fast (SPP) (He et al., 2015), the new version uses Fast-SPP (SPPF) to improve the processing speed of feature information. And replace all C3 modules in the backbone network with Pconv reduces the number of parameters as well as the computational effort of the network model, making the model able to run on low performance servers and more suitable for deployment on mobile devices.
- (2) Neck: This part mainly consists of Feature Pyramid Networks (Lin et al., 2017) and Perceptual Adversarial Network (Liu et al., 2018), which first fuses the input insulator feature maps from top to bottom to transfer the semantic information from the deep layer to the shallow layer to enhance the semantic representation at multiple scales, and then performs a bottom-up feature fusion to transfer the location information from the bottom layer to the deep layer to enhance the localization at multiple scales. The fusion of C3 module with CBAM attention mechanism in this part strengthens the ability of Neck part for fusion of target feature information, especially for small target insulators and self-detonation defective parts of insulators, and also reduces the complexity of the model to some extent.
- (3) Head: This part mainly detects 3 different scales, including some convolutional layers, pooling layers and fully connected layers, etc. Its role is to perform multi-scale target detection on the feature maps extracted from the backbone network. The model proposed in this article uses WIoU loss to improve detection accuracy and convergence speed in this section.

In order to verify the effectiveness of the method in this paper, the heat map visualization operation (Quan et al., 2022) was performed on the insulator feature extraction process in complex backgrounds, as shown in Figure 6, from which it can be seen that after the convolutional layer extracts the shallow information of insulators, the model can effectively segment the region where the target is located from the background environment; the sampling effect is obviously enhanced after the second stage C3CBAM feature extraction; after the third and fourth stage processing, the higher-level semantic information of the feature map has been more blurred, and the extracted insulator features have become abstracted. From the visualization results of the heat map, it is clear that the algorithm of this paper can more fully extract the color, texture, shape and edge information of insulator defects in the image, so as to quickly and accurately detect defective insulators.

Meanwhile, we designed a visual detection interface based on PyQt5 for the algorithm in this paper, as shown in Figure 7, which mainly has the following functions:

Model, select different models. Input, select the files to be detected, including the detection of pictures and videos in local files, and also has the function of real-time detection using the device's camera and supports RTSP video streaming. The ability to adjust the IoU, confidence level and frame rate delay in the detection process of the model, when reducing the IoU and confidence level, can make the model detect more targets, but the detection error is higher. When IoU and confidence are adjusted up, the accuracy of detection increases and the rate of missed detection increases. The delay can also be selected independently during the detection process. The interface also has the functions of start, pause and end, and the detection results are counted at the bottom left of the interface, and the results are automatically saved when the detection is completed.



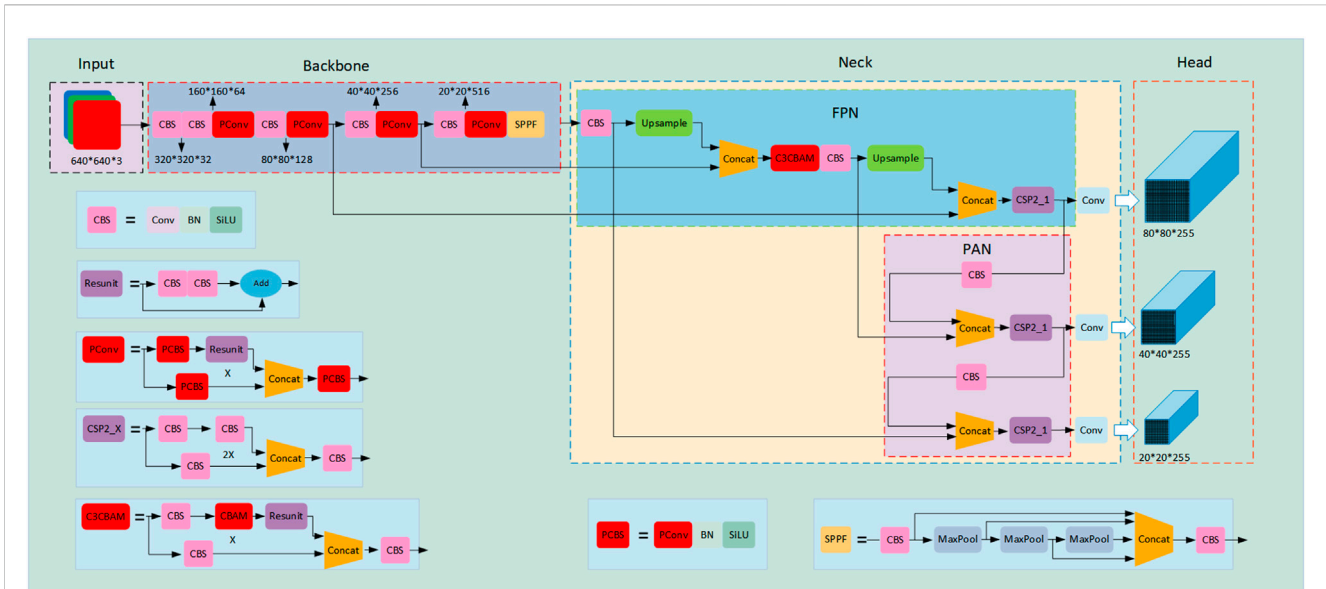


FIGURE 5 FA-YOLOv5 network structure diagram.

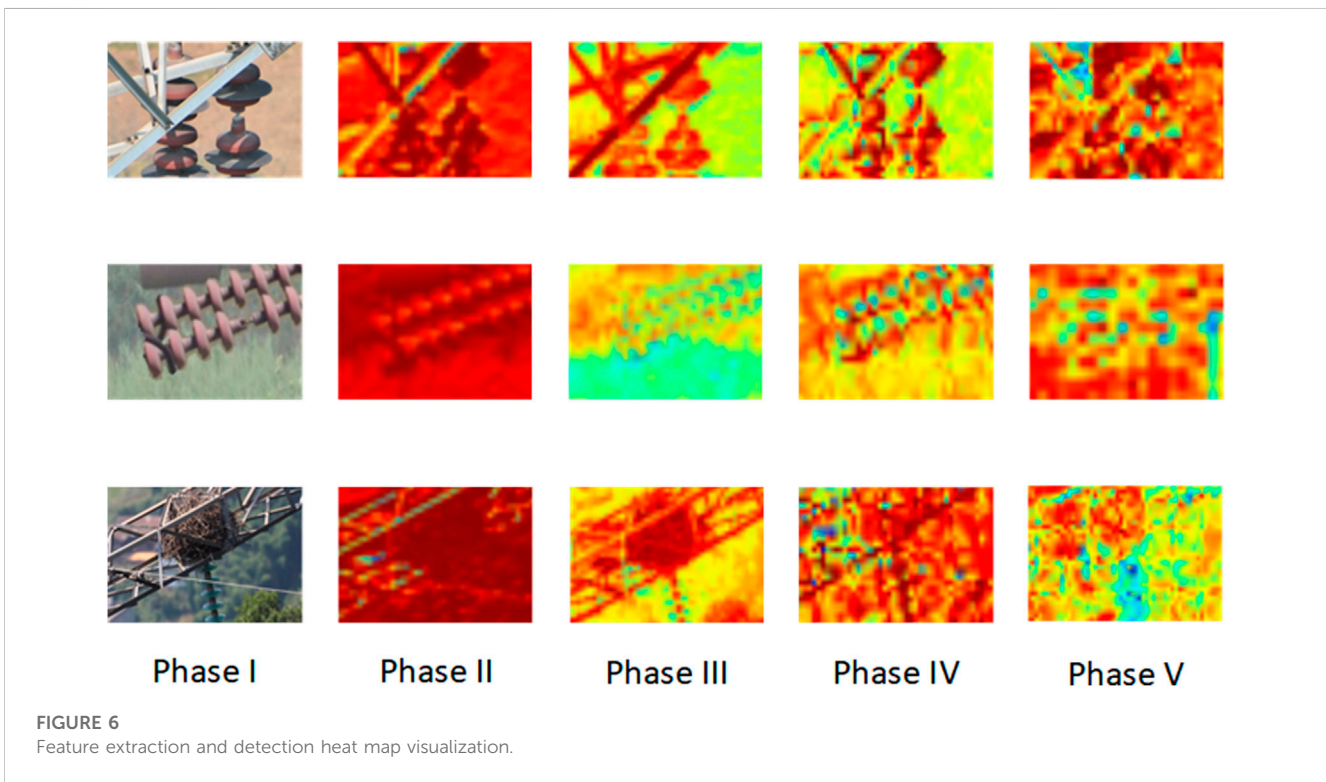


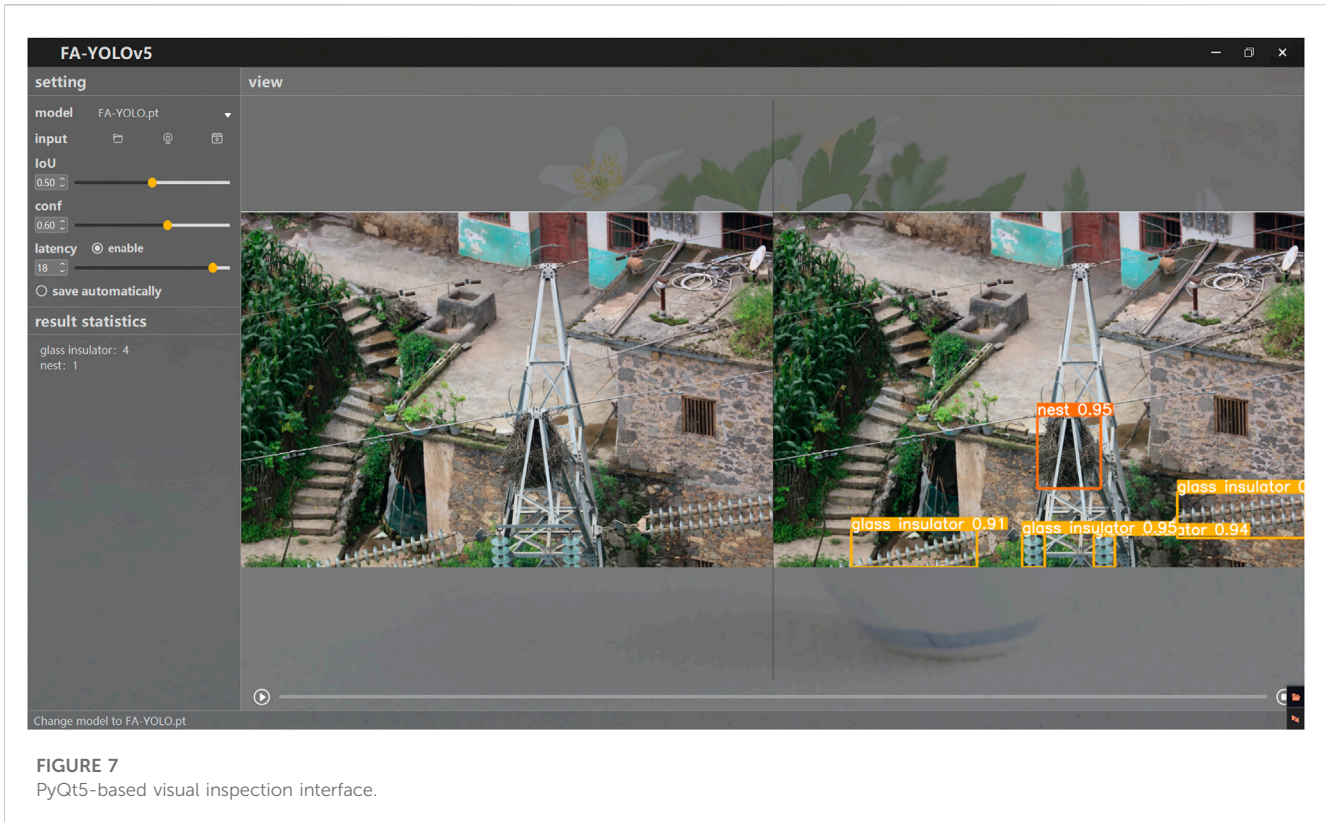
FIGURE 6 Feature extraction and detection heat map visualization.

## 4 Experimental results and analysis

### 4.1 Experimental environment and data pre-processing

The operating system for the experiments in this paper is Windows 11, the CPU model is Intel(R) Core(TM) i7-11700 2.5GHz, 64GB RAM,

and the GPU model is GeForce RTX 3060 Laptop GPU with 12G video memory size of the workstation. The experimental environment is Python 3.8, GPU acceleration software CUDA 11.1 and CUDNN 8.1.0. The datasets used in this paper are mainly derived from three parts, with a total of 1006 insulator images. The first part is the Chinese power line insulator datasets (CPLID) (Raimundo, 2020), which includes 600 images of normal insulators and 248 images of self-exploding



**FIGURE 7**  
PyQt5-based visual inspection interface.

insulators. The second part is 40 self-exploding images of glass insulators disclosed by Baidu Flying Paddle; the third part is 118 images containing self-exploding glass insulators as well as bird's nests taken on site in a southern power grid. The data set was also labeled by LabelImg software, and the labeling labels were divided into: normal insulator (insulator), self-detonation defect (defect), bird's nest (nest) and glass insulator (glass insulator). Due to the lack of sufficient number of datasets, we expanded the number of datasets to 5174 by Gaussian blurring, cropping, brightness variation, and flipping of the existing datasets, and the results of partial data enhancement are shown in Figure 8. And the ratio of training set, validation set and test set is divided randomly in the form of 8:1:1. The input image size is  $640 \times 640$ , the batch size is 16, the initial learning rate is 0.001, the network parameters are updated using SGD, the learning momentum is 0.937, the weight decay is 0.0005, warmup momentum is 0.8, the translate parameter is set to 0.1, and each training is 100 epochs.

## 4.2 Evaluation metrics

In order to accurately evaluate the performance of the algorithm, Precision (P), Recall (R), Average Precision AP and Mean Average Precision (mAP) are the most commonly used model evaluation metrics in the field of target detection, which are calculated as shown in Eqs 9–12, respectively.

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

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

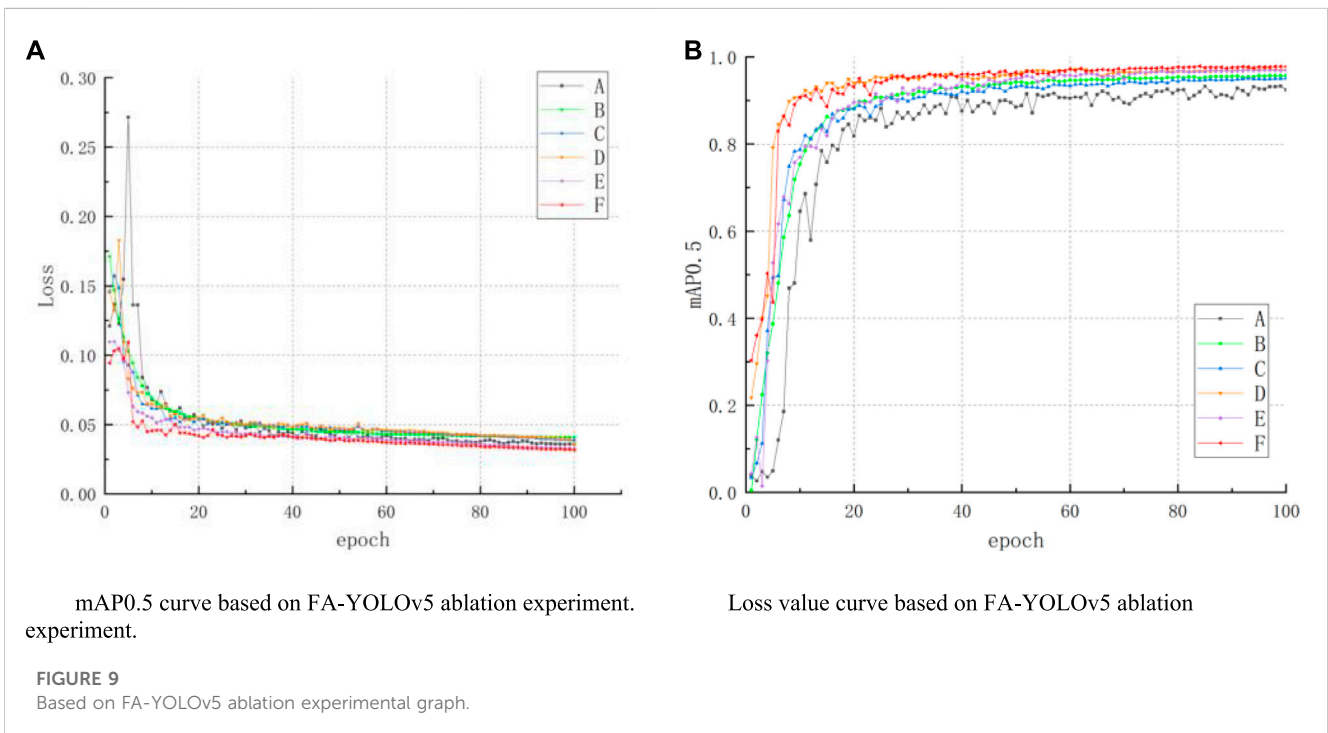
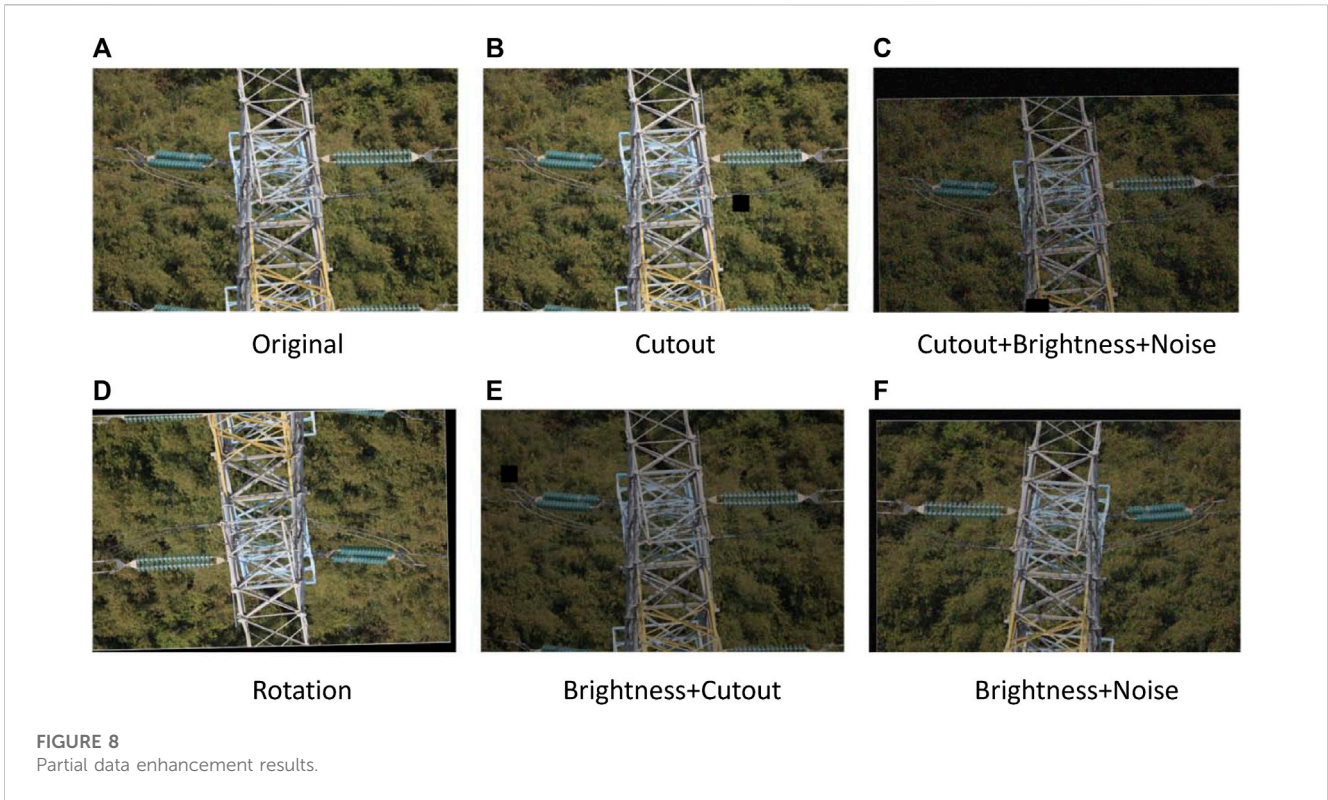
$$\text{AP} = \int_0^1 P(r) dr \quad (11)$$

$$\text{mAP} = \frac{\sum_{i=1}^N AP_i}{N} \quad (12)$$

Where TP denotes the number of positive samples predicted as positive by the model, FP denotes the number of negative samples predicted as positive by the model, i.e., false detection, and FN denotes the number of positive samples predicted as negative by the model, i.e., missed detection. N is the total number of detected categories, and in this paper N is set to 4, i.e., normal insulators, self-detonation defective insulators, bird's nests, and glass insulators. AP is the area enclosed by the PR curve, mAP is the detected average value of AP for each category. The larger the mAP, the better the performance of the algorithm.

## 4.3 Ablation experiment

In order to verify the effectiveness of the algorithm proposed in this paper, mAP, Precision, Recall, parameter quantity, and FPS were used as evaluation indicators to compare the performance of the model through ablation experiments. A total of 6 sets of models were used. Group A is the original datasets for YOLOv5s model training, Group B is the expanded datasets for YOLOv5s model training, Group C, D and E add PConv, C3CBAM and WIoU loss function respectively on the basis of Group B, and Group F (Ours) add PConv, C3CBAM and WIoU loss function on the basis of Group B, and carry out comparative



experiments with the same parameters. The experimental results are shown in Figure 9 and Table 5, respectively.

As can be seen from Table 5, before data enhancement, the YOLOv5 algorithm was not effective in detecting insulator defective parts due to the lack of sufficient defective samples, and after the data enhancement operation, it can be seen that the algorithm has

significantly improved the detection accuracy for all four categories, but there is still some room for improvement in the number of parameters, computation and overall performance of the model. When we use PConv to improve the backbone network part of the model, the number of parameters of the model is reduced by 27.2% and the computation is reduced by 34.4%, while the speed of



TABLE 5 Ablation experiments based on improved FA-YOLOv5.

/	Model	Precision (%)				Recall (%)	mAP0.5 (%)	Parameters(M)	GFLOPs	FPS	Weights (MB)
		Insulator	Defect	Nest	Glass insulator						
A	YOLOv5(Original datasets)	90.2	94.8	87.8	96.7	85.1	92.4	7.0	16.0	81	13.7
B	YOLOv5(Data Enhancements)	94.6	95.5	95.0	97.2	94.5	95.4	7.0	16.0	81	13.7
C	YOLOv5(Data Enhancements + PConv)	93.8	94.7	94.2	96.7	94.0	94.4	5.1	10.7	89	10.0
D	YOLOv5(Data Enhancements + C3CBAM)	95.9	96.3	96.8	96.4	96.1	96.5	6.9	15.7	80	13.6
E	YOLOv5(Data Enhancements + WIoU)	95.6	97.2	97.5	96.9	96.5	95.9	7.0	16.0	85	13.7
F	YOLOv5(Data Enhancements + PConv + C3CBAM + WIoU)	95.4	97.6	97.2	97.3	96.7	97.1	<b>5.0</b>	<b>10.3</b>	<b>89</b>	<b>9.83</b>

The bold portion of the table indicates the value with the best performance in the metric.

TABLE 6 Comparison of experimental results of different algorithmic models.

Methods	Precision (%)	Recall (%)	mAP0.5 (%)	Parameters(M)	GFLOPs
YOLOv5s	95.5	94.5	95.4	7.0	16.0
YOLOv7	96.9	96.0	96.1	37.2	105.2
YOLOv8s	97.3	<b>96.8</b>	96.6	11.1	28.7
SSD-VGG	85.6	67.8	84.3	26.3	62.7
Faster R-CNN	82.1	70.1	81.6	137.1	370.2
FA-YOLOv5	97.6	96.7	97.1	5.0	10.3

The bold portion of the table indicates the value with the best performance in the metric.

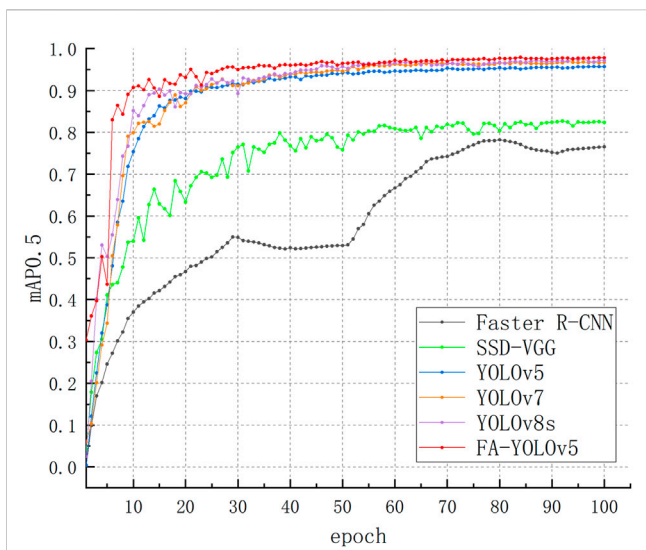
detection is improved to some extent, but the accuracy of detection is reduced. The fusion of the CBAM attention mechanism with the C3 module not only improves the detection accuracy but also reduces the complexity of the network. To further improve the performance of the model, we use Wise\_Loss as the loss function of the model. Finally, a faster and more accurate model FA-YOLOv5 is proposed, which has 1.6% higher mAP value, 28.6% lower number of parameters, and 35.7% lower computational effort compared to the original model.

The comparison graph of experimental results is shown in Figure 9. Analysis of the mAP0.5 graph in Figure 9A shows that the convergence of the original algorithm is slow and the accuracy is low when no data augmentation is performed. After the data enhancement of the defective samples, the situation is significantly improved, and it can also be seen that the algorithm of this paper has stabilized at the 40th round and achieved a high detection accuracy. From the Loss plot in Figure 9B, it can be seen that the loss value of the algorithm for training the original datasets only starts to

stabilize in the 53rd epoch, while the loss value of the algorithm after doing data enhancement operation on the original datasets slowly stabilizes after 30 epochs of training, but the loss values of the algorithm are all improved, and it can be seen from the curve sets of Group E (WIoU) and Group F (Ours) that the improvement of the loss function in this paper has obvious effect on speeding up the convergence, while the loss value reaches the minimum.

#### 4.4 Comparison experiments

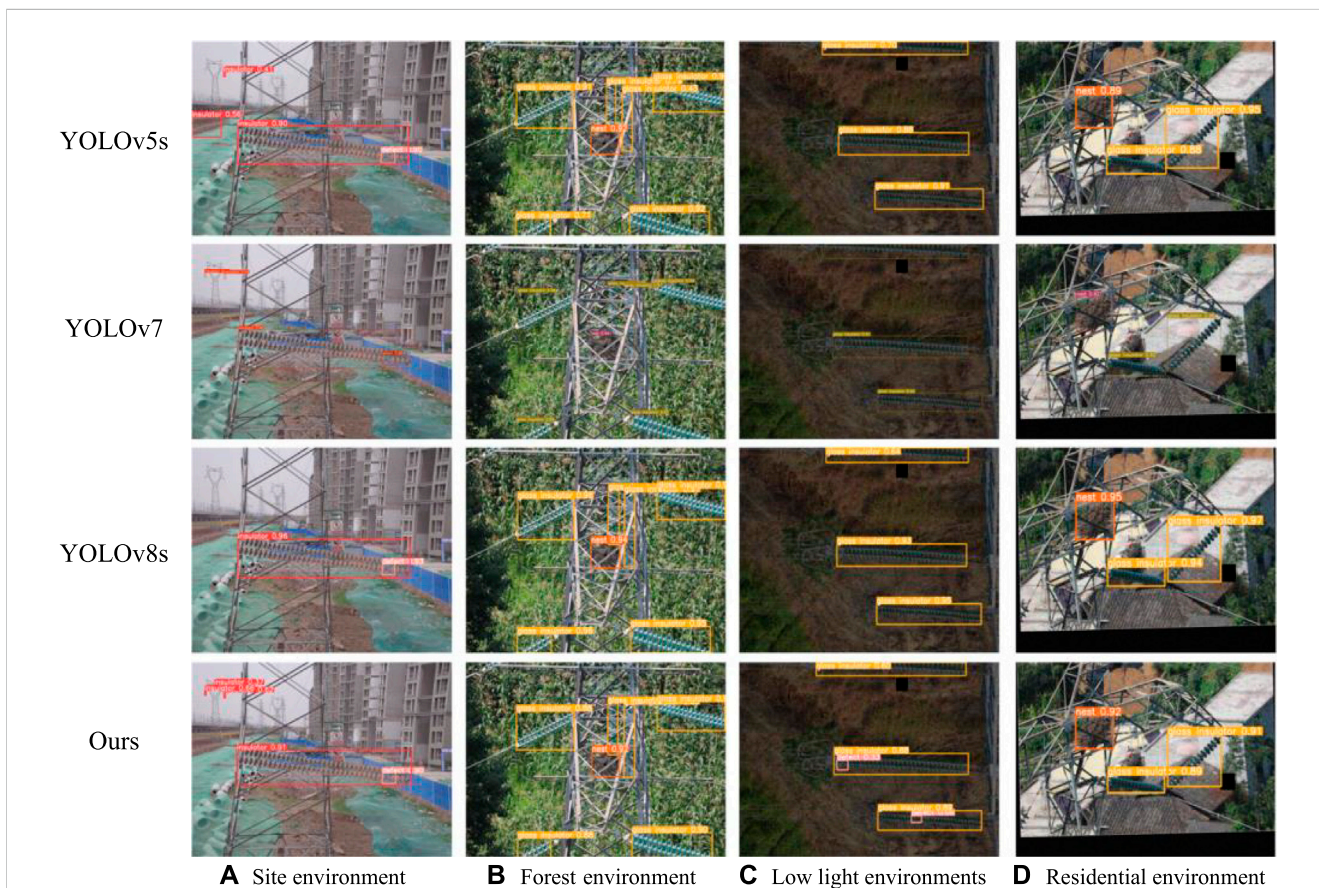
To further verify the superiority and feasibility of the algorithm in this paper, we conducted comparison experiments on the unimproved YOLOv7, YOLOv8s, SSD and Faster R-CNN algorithms with optimal parameters, and the datasets used for the experiments were all self-built insulator defect datasets in this paper, and the precision, recall and average precision during the experiments of the mean value are shown in Table 6.



**FIGURE 10**  
Comparison of experimental results of mAP0.5 curves for different models.

accuracy, but compared with YOLOv5s, their number of parameters and computation amount are larger. Faster R-CNN, as a typical Two-stage algorithm, has the highest number of parameters and the largest computation amount, and also the worst performance among all the compared algorithm models. SSD algorithm, as one of the typical One-stage algorithms, has only a little bit more parameter and computation amount than the YOLOv5 algorithm, but due to the fact that the last layer of the feature map of the network structure is too small, it is easy to lose the feature information of the target, which leads to the loss of feature information of the target. which leads to easy loss of the target's feature information, so the detection effect for this dataset is also poor. In order to balance the detection accuracy and model complexity, this paper proposes a lightweight model FA-YOLOv5 with better detection performance on the basis of YOLOv5 model, which has the highest detection accuracy and the least network parameters and computation among the listed models, and it is more suitable for deploying on mobile devices for transmission line inspection such as UAVs, which proves the feasibility of the method in this paper. Meanwhile, from the mAP0.5 curve graph in Figure 10, it can also be more intuitively seen that the algorithm proposed in this paper has a better convergence speed, and at the same time, it also has a better detection accuracy, and its detection performance is better than that of other comparative algorithms, which further proves the effectiveness of the algorithm in this paper.

From Table 6, it can be seen that among the unimproved algorithm models, YOLOv8s and YOLOv7 models, as the latest target detection algorithms nowadays, have high detection



**FIGURE 11**  
Comparison of detection results of different algorithms.



Finally, this paper compares the detection result graphs of the four models with the highest mAP values, and the comparison results are shown in Figure 11. From Figures 11B,D, it can be seen that the YOLOv7 and YOLOv8s algorithms have high detection accuracy for glass insulators, but the detection of the small targets as well as defective regions is not effective, and there are serious leakage cases.

From Figure C, it can be seen that YOLOv5s, YOLOv7 and YOLOv8s algorithms have lower detection accuracy under the interference of low-light as well as Gaussian noise, and the model's anti-interference ability is weaker, and the robustness is insufficient. In contrast, the FA-YOLOv5 proposed in this paper can accurately detect the small target insulators in the distance as well as the occluded insulators, and can accurately detect the insulators and their defective regions even under low light, and at the same time, it also has high detection accuracy under the interference of Gaussian noise and good anti-interference ability, which further proves that this paper's method can be applied to the presence of small targets under complex backgrounds in the presence of occlusion as well as dense and other cases can have a better detection effect.

## 5 Conclusion

In view of the slow detection speed and low accuracy, even leakage detection and false detection caused by the current insulator defect detection model with large number of parameters and large computation, as well as the complex environment in which insulators are located, the small percentage of defective parts, and the existence of mutual occlusion between insulators, this paper improves the YOLOv5s algorithm and proposes a lightweight FA-YOLOv5s algorithm based on it, with the following main contributions.

- 1) Strengthening feature fusion: By integrating the CBAM attention mechanism into the C3 module, the characteristics of both the attention mechanism and the C3 module are combined to enhance the algorithm's ability to fuse target feature information. This allows the feature information to better propagate to the detection head, resulting in improved detection accuracy.
- 2) Lightweight processing: Lightweight improvement is made to the convolutional modules in the main network of the model, balancing the relationship between network structure complexity and detection performance, so that the network reduces the number of parameters and computations at the cost of a small decrease in accuracy.
- 3) In this paper, the CIoU loss function used in the original model is improved to a WIoU loss function, which balances the variability in sample quality and improves the overlap between the prediction frame and the bounding box to improve the accuracy of the detection compared to CIoU.
- 4) A visualized software interface for defective insulator detection is designed, which enables a more intuitive observation of the detection results of the model.

However, during the experimental process, we found that the insulator defective dataset used in this paper is of a single type, and the data enhancement method can only expand the number of samples, and cannot enrich the diversity of the background environment, resulting in limited application in real scenarios. In the next work, we will consider going to the field to actually shoot more insulator images in different scenes, to further improve the robustness of the algorithm and the diversity of the dataset, and optimize the effect of YOLOv5s algorithm on the detection of defective insulators.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary materials, further inquiries can be directed to the corresponding author.

## Author contributions

JL: Funding acquisition, Resources, Writing–review and editing. MH: Data curation, Methodology, Writing–original draft. JD: Investigation, Validation, Writing–review and editing. XL: Conceptualization, Resources, Writing–review and editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported in part by the National Natural Science Foundation of China (62176067); Joint Fund for Basic and Applied Basic Research in Guangdong Province (20A1515111162); Scientific and Technological Planning Project of Guangzhou (201903010041, 202103000040); Key Project of Guangdong Province Basic Research Foundation (2020B1515120095); Project Supported by Guangdong Province Universities and Colleges Pearl River Scholar Funded Scheme (2019).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M. (2020). *Yolov4: optimal speed and accuracy of object detection*. arXiv preprint arXiv:2004.10934, Available at: <https://doi.org/10.48550/arXiv.2004.10934>.
- Chen, B. (2020). Fault statistics and analysis of 220-kv and above transmission lines in A southern coastal provincial power Grid of China. *IEEE Open Access J. Power Energy 7*, 122–129. doi:10.1109/oajpe.2020.2975665
- Chen, J., Kao, S.-h., He, H., Zhuo, W., Wen, S., Lee, C.-H., et al. (2023). Run, don't walk: chasing higher flops for faster neural networks. arXiv preprint arXiv:2303.03667, Available at: <https://doi.org/10.48550/arXiv.2303.03667>.
- Chen, Y., Liu, H., Chen, J., Hu, J., and Zheng, E. (2023). Insu-YOLO: an insulator defect detection algorithm based on multiscale feature fusion. *Electronics* 12 (15), 3210. doi:10.3390/electronics12153210
- El-Hag, A. (2021). Application of machine learning in outdoor insulators condition monitoring and diagnostics. *IEEE Instrum. Meas. Mag.* 24, 101–108. doi:10.1109/mim.2021.9400959
- Everingham, M., Eslami, S. A., Van Gool, L., Williams, C. K., Winn, J., and Zisserman, A. (2015). The pascal visual object classes challenge: a retrospective. *Int. J. Comput. Vis.* 111, 98–136. doi:10.1007/s11263-014-0733-5
- Gao, Z., Yang, G., Li, E., and Liang, Z. (2021). Novel feature fusion module-based detector for small insulator defect detection. *IEEE Sensors J.* 21, 16807–16814. doi:10.1109/jsen.2021.3073422
- Han, G., He, M., Gao, M., Yu, J., Liu, K., and Qin, L. (2022). Insulator breakage detection based on improved Yolov5. *Sustainability* 14, 6066. doi:10.3390/su14106066
- He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). “Mask R-CNN,” in Proceedings of the IEEE international conference on computer vision, Venice, Italy, 22–29 October 2017, 2961–2969.
- He, K., Zhang, X., Ren, S., and Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. pattern analysis Mach. Intell.* 37, 1904–1916. doi:10.1109/tpami.2015.2389824
- Lewy, D., and Mańdziuk, J. (2023). An overview of mixing augmentation methods and augmentation strategies. *Artif. Intell. Rev.* 56, 2111–2169. doi:10.1007/s10462-022-10227-z
- Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). “Feature pyramid networks for object detection,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2117–2125.
- Lin, T. Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., et al. (2014). “Microsoft coco: common objects in context,” in Proceedings, Part V 13 Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, 740–755. Springer.
- Liu, S., Qi, L., Qin, H., Shi, J., and Jia, J. (2018). “Path aggregation network for instance segmentation,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 8759–8768.
- Lu, H., Li, Y., Mu, S., Wang, D., Kim, H., and Serikawa, S. (2017). Motor anomaly detection for unmanned aerial vehicles using reinforcement learning. *IEEE internet things J.* 5, 2315–2322. doi:10.1109/ijot.2017.2737479
- Miao, X., Liu, X., Chen, J., Zhuang, S., Fan, J., and Jiang, H. (2019). Insulator detection in aerial images for transmission line inspection using single shot multibox detector. *IEEE Access* 7, 9945–9956. doi:10.1109/access.2019.2891123
- Quan, Y., Zhang, D., Zhang, L., and Tang, J. (2022). Centralized feature pyramid for object detection. arXiv preprint arXiv:2210.02093, Available at: <https://doi.org/10.48550/arXiv.2210.02093>.
- Raimundo, A. (2020). *Insulator data set-Chinese power line insulator dataset (cplid)*. IEEE Dataport.
- Redmon, J., and Farhadi, A. (2018). Yolov3: an incremental improvement. arXiv preprint arXiv:1804.02767, Available at: <https://doi.org/10.48550/arXiv.1804.02767>.
- Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster R-CNN: towards real-time object detection with region proposal networks. *Adv. neural Inf. Process. Syst.* 28, 1137–1149. doi:10.1109/TPAMI.2016.2577031
- Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., and Savarese, S. (2019). “Generalized intersection over union: a metric and A loss for bounding box regression,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 658–666.
- Shuang, F., Wei, S., Li, Y., Gu, X., and Lu, Z. (2023). Detail R-CNN: insulator detection based on detail feature enhancement and metric learning. *IEEE Trans. Instrum. Meas.* 72, 1–14. Art no. 2524414. doi:10.1109/tim.2023.3305667
- Tan, P., Li, X. f., Ding, J., Cui, Z. s., Ma, J. e., Sun, Y. l., et al. (2022). Mask R-CNN and multifeature clustering model for catenary insulator recognition and defect detection. *J. Zhejiang University-SCIENCE A* 23, 745–756. doi:10.1631/jzus.a2100494
- Tong, Z., Chen, Y., Xu, Z., and Yu, R. (2023). Wise-iou: bounding box regression loss with dynamic focusing mechanism. arXiv preprint arXiv:2301.10051, Available at: <https://doi.org/10.48550/arXiv.2301.10051>.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). “Attention is all you need,” in *Advances in neural information processing systems* 30 (NeurIPS Proceedings).
- Wang, J., Zhang, T., Xue, X., and Chen, L. (2022). “Real-time recognition of transmission line insulators under complex backgrounds: a Yolov5s approach,” in 2022 4th International Conference on Power and Energy Technology (ICPET), Beijing, China, 28–31 July 2022, 77–83.
- Wei, L., Angelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., et al. (2016). “Ssd: single shot multibox detector,” in Proceedings, Part I 14. Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016. Springer International Publishing.
- Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. (2018). “Cbam: convolutional Block attention module,” in Proceedings of the European conference on computer vision (ECCV), 3–19.
- Yang, W., Liu, Y., and Li, Y. (2021). “Insulator UV image fault detection based on deep learning,” in 2021 IEEE International Conference on Electrical Engineering and Mechatronics Technology (ICEEMT) (IEEE), Qingdao, China, 02–04 July 2021, 632–635.
- Yang, Z., Xu, Z., and Wang, Y. (2022). Bidirection-Fusion-Yolov3: an improved method for insulator defect detection using uav image. *IEEE Trans. Instrum. Meas.* 71, 1–8. doi:10.1109/tim.2022.3201499
- Yi, W., Ma, S., and Li, R. (2023). Insulator and defect detection model based on improved yolo-S. *IEEE Access* 11, 93215–93226. doi:10.1109/access.2023.3309693
- Yu, J., Jiang, Y., Wang, Z., Cao, Z., and Huang, T. (2016). “Unitbox: an advanced object detection network,” in Proceedings of the 24th ACM international conference on Multimedia, Amsterdam, The Netherlands, October 15 - 19, 2016, 516–520.
- Yu, Y., Cao, H., Wang, Z., Li, Y., Li, K., and Xie, S. (2019). Texture-and-shape based active contour model for insulator segmentation. *IEEE Access* 7, 78706–78714. doi:10.1109/access.2019.2922257
- Zhang, L.-P., Zhao, J.-M., and Ren, Y.-F. (2018). “Research on multiple features extraction technology of insulator images,” in 2018 10th International Conference on Modelling, Identification and Control (ICMIC) (IEEE), Guiyang, China, 02–04 July 2018, 1–6.
- Zhang, S., Qu, C., Ru, C., Wang, X., and Li, Z. (2023). Multi-objects recognition and self-explosion defect detection method for insulators based on lightweight GhostNet-YOLOV4 model deployed onboard UAV. *IEEE Access* 11, 39713–39725. doi:10.1109/access.2023.3268708
- Zhang, Y.-F., Ren, W., Zhang, Z., Jia, Z., Wang, L., and Tan, T. (2022). Focal and efficient iou loss for accurate bounding box regression. *Neurocomputing* 506, 146–157. doi:10.1016/j.neucom.2022.07.042
- Zhao, W., Xu, M., Cheng, X., and Zhao, Z. (2021). An insulator in transmission lines recognition and fault detection model based on improved faster RCNN. *IEEE Trans. Instrum. Meas.* 70, 1–8. Art no. 5016408. doi:10.1109/tim.2021.3112227
- Zheng, H., Sun, Y., Liu, X., Djike, C. L. T., Li, J., Liu, Y., et al. (2020a). Infrared image detection of substation insulators using an improved fusion single shot multibox detector. *IEEE Trans. Power Deliv.* 36, 3351–3359. doi:10.1109/tpwrd.2020.3038880
- Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., and Ren, D. (2020b). Distance-iou loss: faster and better learning for bounding box regression. *Proc. AAAI Conf. Artif. Intell.* 34, 12993–13000. doi:10.1609/aaai.v34i07.6999
- Zheng, Z., Wang, P., Ren, D., Liu, W., Ye, R., Hu, Q., et al. (2021). Enhancing geometric factors in model learning and inference for object detection and instance segmentation. *IEEE Trans. Cybern.* 52, 8574–8586. doi:10.1109/tycb.2021.3095305