



High-Voltage Transmission Line Foreign Object and Power Component Defect Detection Based on Improved YOLOv5

Shanshan Wang^{1,2} · Weiwei Tan¹ · Tengfei Yang¹ · Liang Zeng^{1,2} · Wenguang Hou³ · Quan Zhou³

Received: 21 April 2023 / Revised: 6 August 2023 / Accepted: 15 August 2023 / Published online: 15 September 2023
© The Author(s) under exclusive licence to The Korean Institute of Electrical Engineers 2023

Abstract

With the outstanding performance of deep learning in the field of computer vision, the automatic visual detection of foreign bodies in transmission lines and electrical equipment defects by inspection robots and Unmanned Aerial Vehicles based on neural networks has become an attractive topic in smart grid. However, in practical application scenarios, small-sized target defects pose a great challenge to the detection accuracy of existing mainstream deep learning detection networks with limited perceptual fields. To solve the above problems, the paper proposes a detection model of YOLOv5 transmission line inspection image. Firstly, the key target images under different backgrounds and attitudes are collected and preprocessed. Specifically, in order to improve the perception ability of networks for small-sized targets, the K-means clustering algorithm is used to optimize the size of the anchor box, which effectively improves the fitting ability of the key target features. Then, recursive gated convolution is used as the backbone network to improve the ability to extract key target features. Finally, considering the concealment of small-scale features, the space-to-depth convolution module is added to the neck network to realize down-sampling and retain all the information in the channel dimension. In addition, a feature prediction layer is added to optimize the scale of network detection, and a Simple Parameter-Free Attention Module is added to further optimize the characterization of network features. The experimental results show that the accuracy and recall of the proposed network are 96.8% and 93.3%, respectively. The average detection accuracy reaches 97.1%, which is 3.8% higher than that of YOLOv5 network, 5.2% higher than YOLOv6 and 1.0% higher than that of YOLOv7 network. The proposed method significantly improves the detection performance of critical targets and defects of high-voltage transmission lines.

Keywords Transmission line inspection · Aerial images · Object detection · Recursive gated convolution

1 Introduction

Weiwei Tan, Tengfei Yang, Liang Zeng, Wenguang Hou and Quan Zhou have contributed equally to this work.

✉ Quan Zhou
zhouquan910@hust.edu.cn

Shanshan Wang
wangshanshan@hbust.edu.cn

¹ School of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan 430068, Hubei Province, China

² Hubei Engineering Research Center for Safety Monitoring of New Energy and Power Grid Equipment, Hubei University of Technology, Wuhan 430068, Hubei Province, China

³ College of lift Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, Hubei Province, China

Insulators and dampers, as essential power components in transmission lines, play an critical part of supporting the line, blocking the current, and preventing damage to the transmission line [1]. They are subject to prolonged exposure to the outside circumstances, often corroded by storms, rain, and snow, inevitably resulting in damage and causing defects. In addition, bird nest on transmission lines can lead to discharging and tripping of lines. These phenomena seriously harm the safe operation of high-voltage lines and decrease the longevity of electrical equipment [2]. Consequently, precisely identifying and positioning the defects and foreign bodies on the electrical equipment on the transmission line has a profound significance for ensuring the healthy operation of the power system [3].

Traditional transmission line inspection often relies on manual inspection, but high-voltage transmission lines cover

a wide range and have complex terrain. It requires staff to observe closely, not only does it consume a considerable amount of manpower and material resources, but the efficiency and safety are also very low. Recently, with the progress of UAV technology and the increasing range of applications, the transmission line images collected by UAVs is becoming increasingly popular [4–7].

At present, vision-based methods for detecting the health status of transmission lines are gradually emerging. It mainly uses image processing methods to identify and locate foreign objects and electrical equipment and defects on transmission lines to achieve daily inspection of transmission lines. In addition, compared with other types of detection objects, transmission line defect detection has the characteristics of large target differences and complex backgrounds.

The traditional target detection algorithm is usually accomplished by two relatively independent processes of feature extraction and feature recognition. Image feature extraction is mainly based on manual feature extraction, which depends on the size, shape, texture and other features of the target. Feature recognition methods mainly include histogram of oriented gradient (HOG) [8], scale-invariant feature transform (SIFT) [9] and so on. For instance, two common hough transform techniques and a line tracing algorithm are used in [10] to realize the detection of power line. In [11], a methodology for classifying insulator damage conditions is proposed through the matching and fusion of contour features and gray similarity. However, traditional methods have low computational complexity and slow processing speed, and are easily affected by complex background and illumination conditions.

Compared with traditional methods, the deep learning object detection methods represented by convolutional neural network can automatically learn image features that are difficult for human observation from a great deal of image, and are applicable to multiple types of scenes at the same time. In target detection, there are two main types of neural networks that are commonly used. The first type is a two-stage model that relies on region extraction, such as region-convolutional neural networks (Faster R-CNN) [12]. The second type is a single-stage model that uses regression, such as you only look once (YOLO) series [13], single shot detector (SSD) [14], etc. These two types of models have been widely used in target detection. Using feature pyramids, an enhanced Faster R-CNN model is proposed in [15], which uses straight line detection, projection and other methods to complete insulator fault detection with complex background. In [16], an improved lightweight YOLOv4 model is proposed to realize insulator detection combined with transfer learning method, achieving 97.26% accuracy. By optimizing the detection network of YOLOv4 model, the extraction of small target features in transmission line inspection images is enhanced in [17]. An enhanced

YOLOv4 insulator defect detection model is proposed in [18], which designed a new backbone network based on residual split-attention networks (ResNest) [19], and adopted multi-scale bidirectional feature pyramid to overcome the obstacle of small-size insulator defects detection. In [20], an image enhancement method based on illumination correction and compensation is proposed to enable real-time monitoring of insulator defects detection using YOLOv5.

The bulk of these methods mentioned above are designed based on the detection task of insulators defective. They have highly personalized characteristics, and it is tough to achieve the expected effect on multiple-category defect detection of small targets. In this case, this paper proposes an improved YOLOv5 transmission line safety detection model for foreign object and power component defect. The primary contributions are as follows:

1. We collect and process a dataset named NID dataset suitable for transmission line inspection, which includes bird nests, defective insulators and dampers.
2. By introducing recursive gated convolution, the extraction ability of target defects in detected images is improved, and the traditional sampling process is replaced by slicing operations to reduce information loss. In addition, a new prediction layer is added to the target detection network to enrich context information and further optimize the detection accuracy of the network.
3. Experiments show that the model can effectively improve the test accuracy of the key goals and defects in the transmission line in complex contexts to achieve a better identification effect. Apart from this, the generalizability of this method is validated on the CPLID.

The remainder of this paper is arranged as follows. Section 2 describes the relevant work of object detection, including the principle of recursive gated convolution and the elaborate design of the object detection framework. The data set and experimental results are introduced detailly in Sect. 3, and Sect. 4 gives the conclusion.

2 Methods

2.1 Recursive Gated Convolution

Recursive gated convolution, proposed in [21], is mainly composed of standard convolution, linear projection, and elemental multiplication. High-order spatial interaction is performed by gated convolution and recursive design, with input adaptive spatial mixing similar to the self-attention

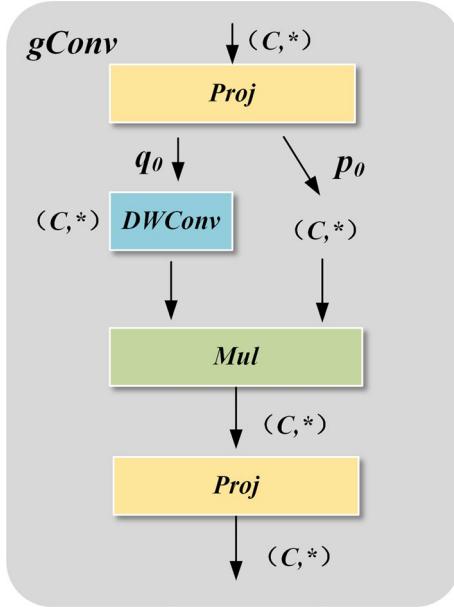


Fig. 1 The structure diagram of gated convolution

mechanism. The core mainly includes gated convolution and cyclic recursive structure design, specifically as follows:

(1) Gated Convolution

Figure 1 shows the structure of gated convolution ($gConv$), which mainly consists of projection layers (Proj), depth-wise convolution layers (DWConv), and multiplication of elements (Mul). C represents the number of channels. $*$ represents other feature information such as the height and width of the feature map.

For example, given an input feature of $x \in R^{HW \times C}$, H and W represent the height and width of the feature map. The output $y = gConv(x)$ of the gated convolution can be expressed as:

$$[p_0^{HW \times C}, q_0^{HW \times C}] = \phi_{in}(x) \in R^{HW \times 2C}, \quad (1)$$

$$p_1 = f(q_0) \odot p_0 \in R^{HW \times C}, y = \phi_{out}(p_1) \in R^{HW \times C} \quad (2)$$

where, ϕ_{in} and ϕ_{out} are the projection layer for performing channel mixing, f represents the depth-wise convolution layer, and \odot refers to the multiplication of elements. The interaction between adjacent features p_0 and q_0 is introduced through element multiplication, so $gConv$ can be regarded as a first-order interaction.

(2) Recursive Gated Convolution

The recursive gated convolution (g^nConv) is designed based on the gated convolution to realize the higher-order interaction of long-range modeling and recursive gating. The details are shown in Fig. 2.

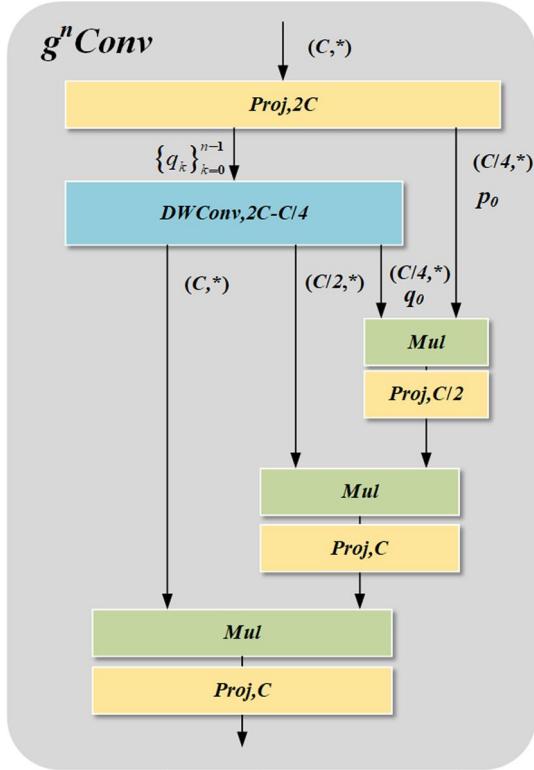


Fig. 2 The structure diagram of recursive gated convolution

In Fig. 2, given the input x , a set of projection features p_0 and $\{q_k\}_{k=0}^{n-1}$ are obtained through the projection layer, the formula can be described as:

$$\left[p_0^{HW \times C_0}, q_0^{HW \times C_0}, \dots, q_{n-1}^{HW \times C_{n-1}} \right] = \phi_{in}(x) \in R^{HW \times (C_0 + \sum_{0 \leq k \leq n-1} C_k)} \quad (3)$$

The gated convolution is recursively executed to gain the output of multiplication of elements at all levels as follows:

$$p_1 = f(q_0) \odot g_0(p_0)/a, \quad k = 0 \quad (4)$$

$$p_2 = f(q_1) \odot g_1(p_1)/a, \quad k = 1 \quad (5)$$

$$\dots \dots \\ p_n = f_{n-1}(q_{n-1}) \odot g_{n-1}(p_{n-1})/\alpha, \quad k = n - 1 \quad (6)$$

In the above formula, f is the 7×7 Depth-wise convolution, and then, the output is proportionally resized to $1/\alpha$ to stabilize the training, and $\{g_k\}$ is used to match the dimensions in different orders, which is specifically expressed as:

$$g_k = \begin{cases} \text{Identity,} & k = 0, \\ \text{Linear}(C_{k-1}, C_k) & 0 \leq k \leq n - 1 \end{cases} \quad (7)$$

The output result p_n of the final recursive iteration is fed into the Proj mapping layer to obtain the outcome of g^nConv .

2.2 Improvement of the Feature Extraction Part

Compared with the two-stage detection network, YOLO network has the benefits of light network structure and rapid detection speed. YOLOv5 is one of the mature YOLO models at present, so the paper chooses YOLOv5 as the backbone of the detection network.

Cross-stage-partial-connections (CSP) Darknet [22] network is used in YOLOv5 to extract features, which achieves the extraction of target feature information through the connection of residual blocks and the stacking of convolutional blocks. However, while accelerating reasoning, the CSP structure ignores the spatial correlation of feature information, which is poor in capturing features of small targets, and there is still scope for improving feature extraction tasks.

In this paper, the feature extraction network is built by alternating connection between convolution batch normalization Sigmoid (CBS) module and feature extraction module named C3CG and spatial pyramid pooling fast (SPPF) structure. CBS module consists of a convolutional layer, a normalization layer, and the activation function. In the detail, C3CG module is designed into two branches, and its structure is shown in Fig. 3. In the figure, branch 1 contains only one convolutional module; branch 2 adopts the four-stage architecture design of transformer, consisting of convolutional modules, layer normalization (Layer Norm), multi-layer perceptron (MLP), and attention modules. Recursive gated convolution (g^nConv) has higher-order spatial interaction and long-range, while ordinary convolution ignores the correlation on the feature space, so the attention module is replaced by g^nConv module to achieve higher-order feature interaction and complete the extraction from shallow features to deep features. In order to further optimize the higher-order feature information, convolutional block attention module (CBAM) [23] is introduced to merge residual information and filter out useless channel and

spatial information. Finally, the two branches are spliced by the concatenate operation, and the higher-order and lower-order feature information is integrated by convolution again to ensure the integrity of the information.

2.3 Design of Feature Fusion Modul

The feature extraction network in the YOLOv5 contains 53 layers in total. When the input image size of the target detection network is $640 \times 640 \times 3$ pixels, the network will sample the original input image by eight times down-sampling, sixteen times down-sampling, thirty-two times down-sampling to get three different sizes of feature graphs, which are $(80, 80, 256)$, $(40, 40, 512)$, $(20, 20, 1024)$. However, these three feature layers are present in the backbone part, situated at the middle, lower middle, and bottom layers, using only this three-layer feature layer as a detection head has low resolution and lacks the underlying feature representation.

To further improve the detection performance of small goals and subtle defects, a 160×160 feature scale module is added to the feature fusion module. By up-sampling the 80×80 feature map and then fusing with the four times down-sampling feature map of the raw image, the output feature scale is expanded from 80×80 to 160×160 . Finally, the obtained multi-scale feature maps are aggregated in the output part for prediction. In addition, the SimAM [24] is introduced into the feature fusion structure attention mechanism to calculate the 3D weights among the features without adding additional parameters, while speeding up the feature fusion. The great details of the improved feature fusion module, which is composed of a top-down feature pyramid network (FPN) [25] and a bottom-up path aggregation network (PAN) [26], are shown in Fig. 4.

The neck network feature fusion module introduces the SPD-Conv [27] layer to replace the ordinary convolutional layer with a step size of 2 to avoid losing too much small target information during the down sampling process. SPD-Conv can reduce redundant pixel information in the difficult task of blurred images or small objects by achieving

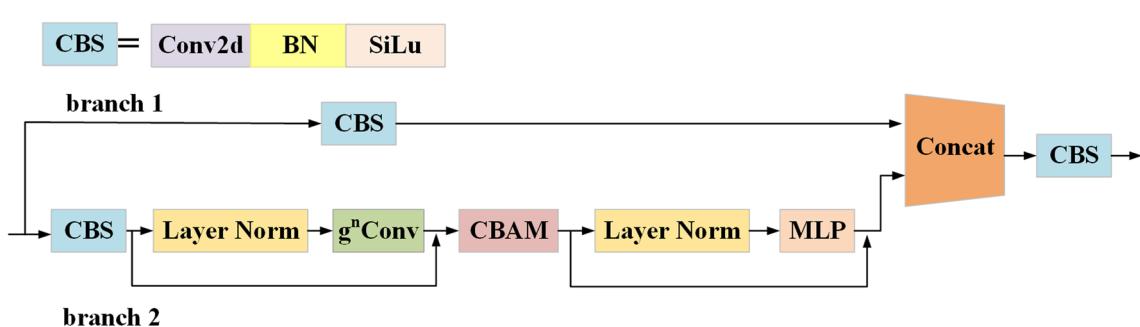


Fig. 3 The structure diagram of C3CG

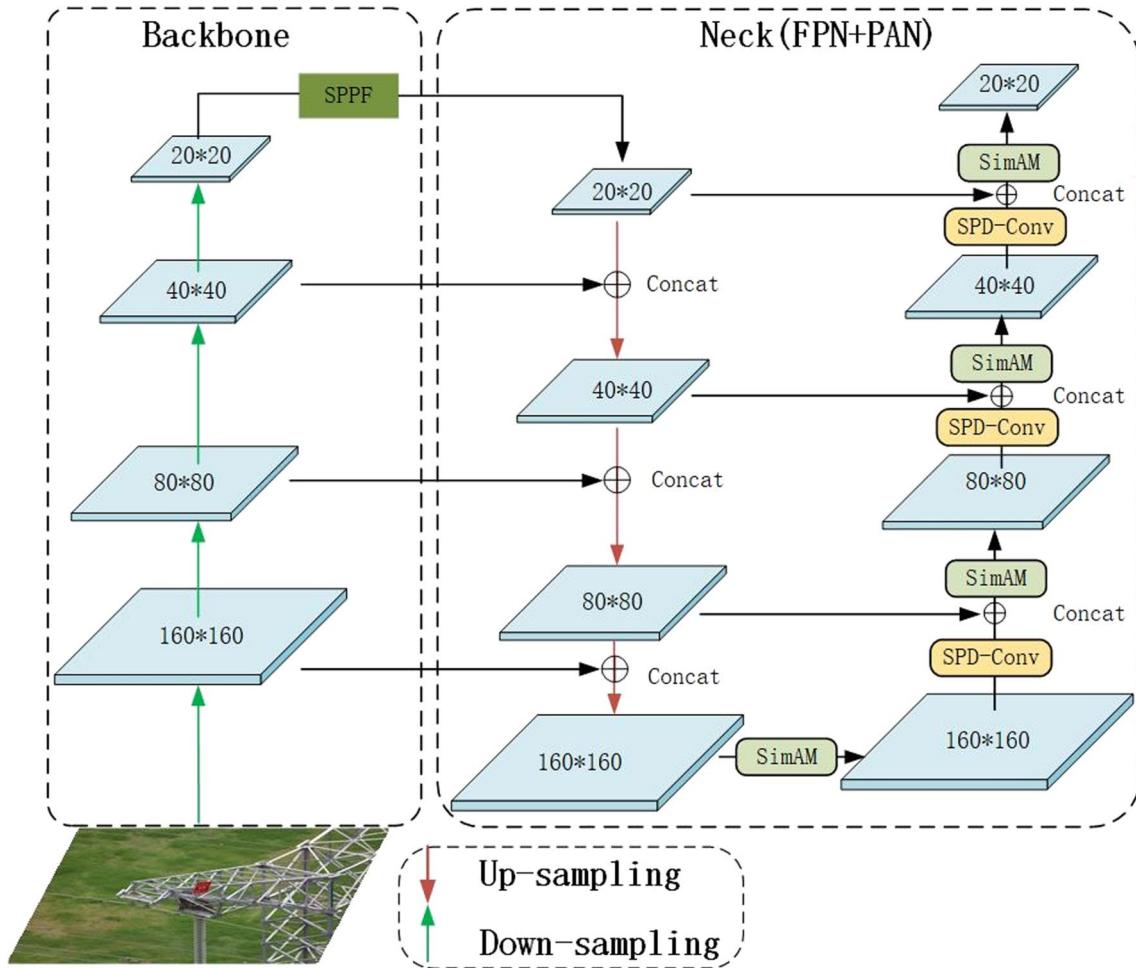


Fig. 4 The structure diagram of FPN and PAN

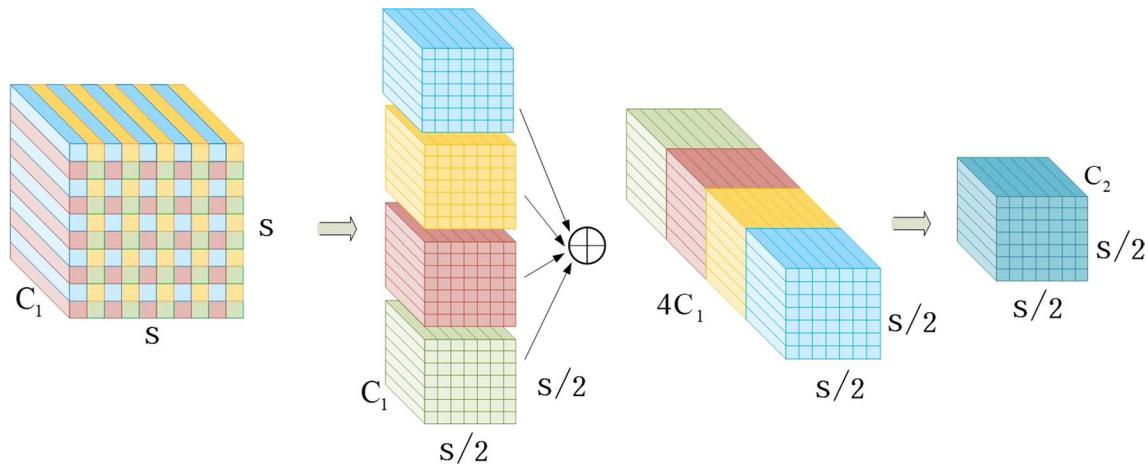


Fig. 5 The structure diagram of SPD-Conv

down-sampling of the feature map through a slice-like operation, as shown in Fig. 5.

Where S represents the height and width of the feature graph, and C represents the channel. An input feature map of size $S \times S \times C_1$ is sliced into 4 sub-feature maps according to an interval sampling unit, and these sub-feature maps are connected along the channel to obtain the intermediate feature map of size $(S/2 \times S/2 \times 4C_1)$. Through a non-stride convolution, the feature map is converted from $(S/2 \times S/2 \times 4C_1)$ to $(S/2 \times S/2 \times C_2)$. The network retains more feature information and enhances the network feature expression capability.

2.4 Improved Target Detection Model

With the improvements presented in the previous two sections, an improved YOLOv5 model is proposed based on the architecture of YOLOv5 target detection network, which improves backbone network, neck network and detection head respectively, the overall structure improvement is revealed in Fig. 6. The 3-channel transmission line inspection images acquired by the UAVs are used as input, and the source images are scaled to 640×640 pixels by adaptive scaling at the network input end. In the improved YOLOv5, C3CG module designed by recursive gated convolution method and the convolution module (CBS) are alternately connected to construct the backbone extraction network,

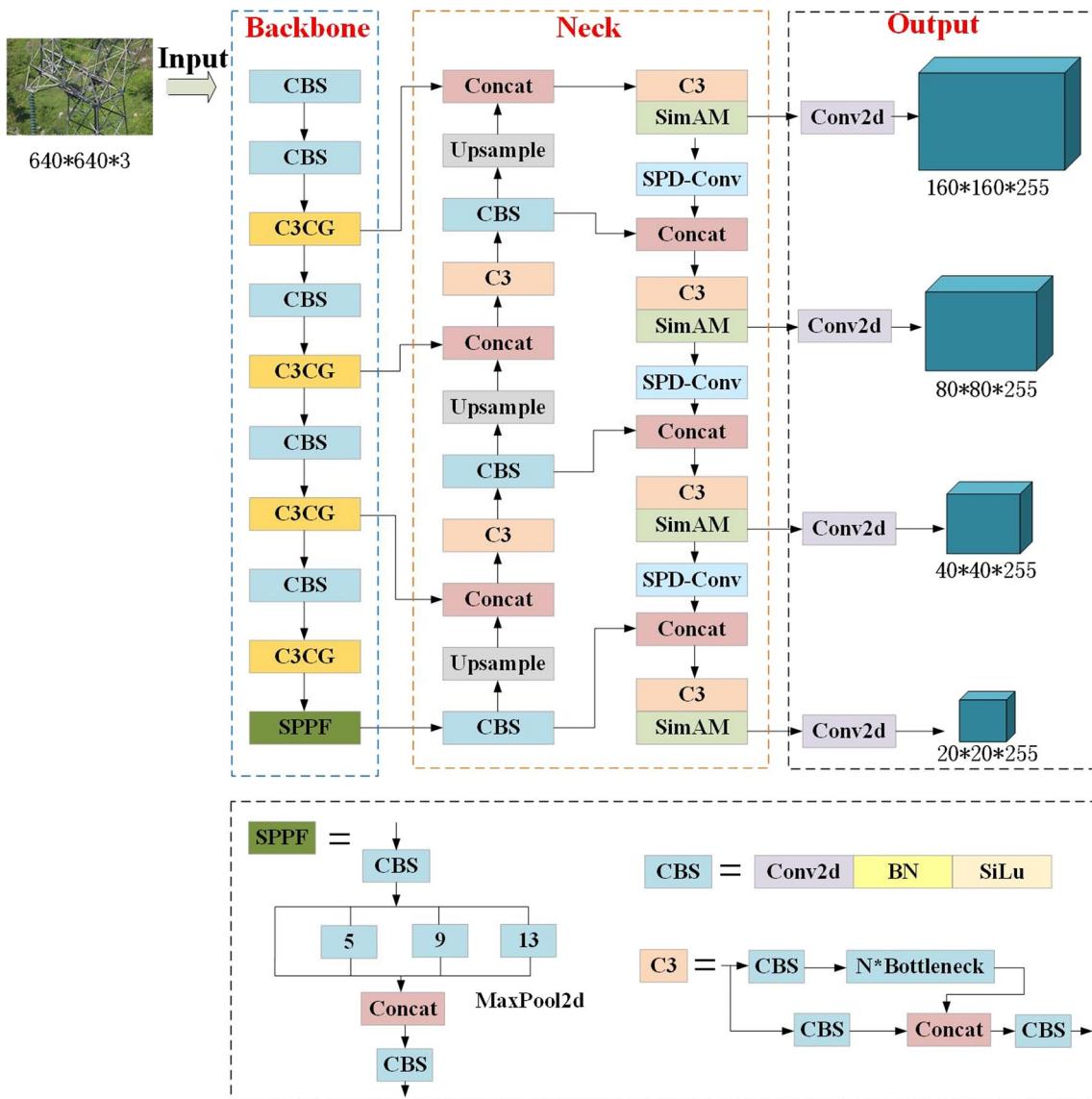


Fig. 6 The structure of foreign object and power component defect detection network for high-voltage transmission lines with improved YOLOv5

which realizes the extraction of more effective feature information through the interaction of higher-order spatial features. After five times of feature down-sampling, the 20×20 feature map is gathered, and the feature map is passed through SPPF to further increase the receptive field of the image. Then, a top-down feature pyramid structure and a bottom-up path aggregation network are used to transfer feature information. In the neck network design, SimAM attention mechanism and 160×160 feature scale are added to enhance shallow information extraction. In particular, sliced convolution SPD-Conv is used to realize down-sampling, avoid the loss of effective feature information, and enhance the detection accuracy of small targets.

3 Experiment Analysis

The hardware configuration of all experimental environments in this paper include Intel(R) Core(TM) i5-12600KF, NVIDIA GeForce RTX 3060 Ti GPU, and the software environment is Windows11 system, Python3.8, PyTorch1.13.0 and CUDA11.3.

In the experimental setup of this study, the initial learning rate is set to 0.01 and is dynamically adjusted using the stochastic gradient descent (SGD) optimizer with momentum size of 0.937 and optimizer weight decay of 0.0005. All experiments are repeated 300 times, with a batch size of 4 and without using pre-trained weights. Additionally, the proposed method in this study uses the same image preprocessing techniques as other comparative algorithms, including Mosaic, random scaling, and horizontal flipping. Moreover, K-means algorithm is used to adjust the anchors of other algorithms.

3.1 Evaluation Metrics

Four metrics, precision (P), recall (R), average precision (AP), and mean average precision (mAP), are used to verify the validity and accuracy of the transmission line multi-target detection system. Higher evaluation metrics indicate better performance, and the formulas for the above measures are shown below, respectively.

$$P = \frac{TP}{TP + FP} \quad (8)$$

$$R = \frac{TP}{TP + FN} \quad (9)$$

$$AP = \int_0^1 P(R)dR \quad (10)$$

$$mAP = \frac{1}{N} \sum_{n=1}^N AP(n) \quad (11)$$

where TP is the number of positive samples identified as positive samples; FP is the number of negative samples misclassified as positive samples; FN is the number of positive samples misclassified as negative samples. AP refers to the area of the P-R curve, The mAP is the average accuracy of all categories; N is the number of detection categories.

3.2 Experimental Dataset

In the current assignment of inspecting transmission line images, there is a lack of publicly available data sets, and some publicly available data sets are of the same quality in terms of detection categories, with relatively single detection objects. For this purpose, combined with the China power line insulator data set (CPLID)[28], a new data set named NID data set is collected and sorted out, some images come from the power supply bureau, which contains 400 images in three categories: bird nest, insulator, and damper. For the purpose of improving identification accuracy, the number of samples is expanded to 1600 images by rotation, translation, and random clipping, which including 1749 bird nest samples, 4961 insulator samples, and 3912 damper samples. A single image contains multiple detection objects. The part of the data sample enhancement is shown in Fig. 7. The NID data set is partitioned randomly into training set, validation set and testing set according to 7:2:1, and the numbers are 1120, 320 and 160, respectively. And the label categories are divided into bird nest (nest), insulator, cracked insulator (break), insulator with defect (defect), damper and damper with defect (defective damper), the details are shown in Table 1.

3.3 Data Analysis

In the data set, there are insulator targets with large aspect ratio, defect targets with small size and dampers, which generally show the characteristics of large difference in targets. The a priori anchor frame parameters based on the common objects in context (COCO) data set would no longer be applicable to this data set, so K-means clustering is used to optimize the a priori anchor frame parameters. The clustering analysis is performed on the a priori frame of the detection objects in the data set, and the clustering results are shown in Fig. 8. Different colors represent different clusters, \times represents the cluster core.

The specific steps of K-means clustering are as follows:

Fig. 7 Example of images with data augmentation

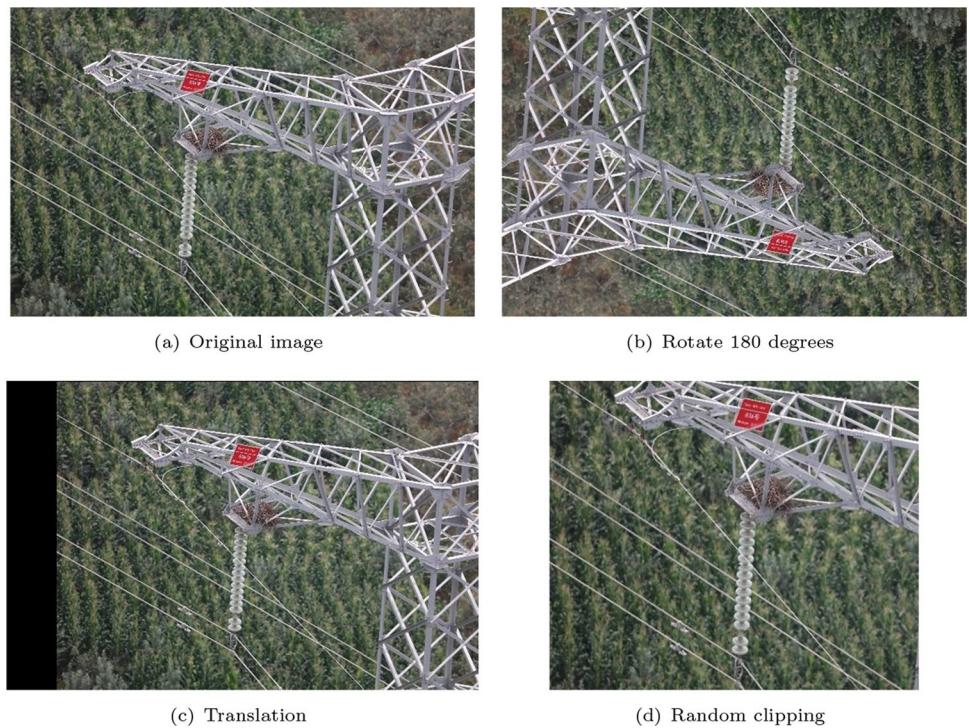


Table 1 Data set partition table

Label category 1	Training set	Validation set	Testing set
Nest	1223	349	177
Insulator	3472	992	497
Break	771	220	111
Defect	336	96	48
Damper	2738	782	392
Defective damper	952	275	134

Table 2 Size distribution of anchor frame after improvement

Cell size/(pixels × pixels)	Anchor size/(pixels × pixels)
160	(92×80),(53×218),(155×91)
80	(21×86),(53×56),(26×185)
40	(23×28),(31×23),(33×38)
20	(15×13),(12×29),(22×18)

1. First, determine the number of clusters k , establish a coordinate system with the width of the anchor frame as the horizontal axis and the height as the vertical axis, and then divide the data set into k clusters;
2. The ratio of the width and height of the anchor frame in each group is randomly selected as the initial centroid of the population;
3. Calculate the distance from the coordinates of all samples to each center of mass, divide them into the nearest clusters, and update the center of mass of each cluster as the expected value of the samples in the cluster;
4. Repeat step 3 until the center of mass of each cluster no longer changes.

In this paper, four prediction scales are used, in the feature diagram, three anchor frames are placed for each target to predict, so k is selected as 12 and the anchor frame

Fig. 8 The graph of K-means clustering results

Table 3 Size distribution of anchor frame after improvement

	C3CG	PAnet-P	SPD-Conv	SimAM	P(%)	R(%)	mAP(%)
Model 1					92.2	85.4	93.3
Model 2	✓				95.5	89.7	94.6
Model 3		✓			93.9	88.6	94
Model 4			✓		94.6	92.1	95.1
Model 5				✓	93.7	91.8	94.2
Model 6			✓	✓	96.2	90.9	96.1
Model 7		✓	✓	✓	97.7	92.6	96.7
Model 8	✓	✓	✓	✓	96.8	93.3	97.1

Bold font indicates that the model performs the best under the same criterion

parameters corresponding to different scales are obtained as shown in Table 2.

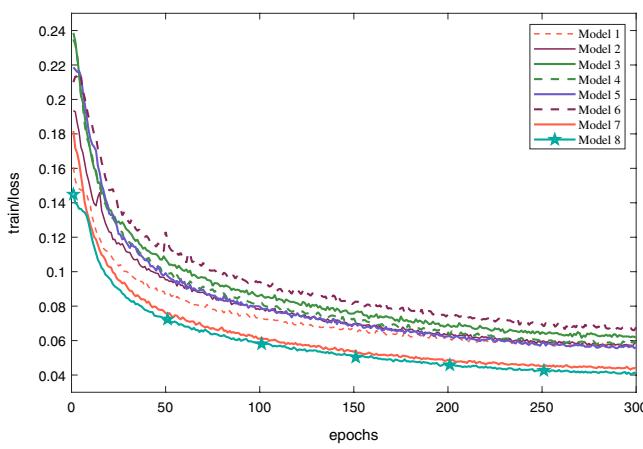
3.4 Results

3.4.1 Effect of Improved Methods

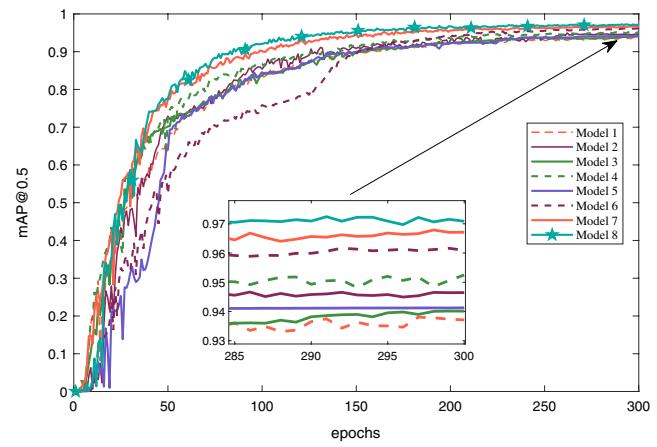
To verify the effectiveness of the improvement measures proposed in the paper, eight models are set up for ablation experiments. The experimental results are shown in Table 3.

In Table 3, The leftmost column represents the network model number, the last three columns represent the evaluation metrics, and the hook symbol in the table represents that the network has incorporated the corresponding improvement strategy. In the table, Model 1 does not incorporate any improvement strategy, and the mAP value is 93.3%. Model 2 incorporates the optimization strategy of constructing a feature extraction network using C3CG modules, resulting in a 1.3% increase in mAP. It has

improved the performance of the network to some extent, that is because the recursive gated convolution pays more attention to the spatial correlation of feature information and can better capture defect information in the detected objects, such as cracked insulators. Model 3 represents the addition of a feature prediction layer, which improves the detection accuracy by 0.7% compared to Model 1. The lower-level features contain more edge information, and adding a new prediction scale can enrich the context information, which is beneficial for object detection to perform position regression. Model 4 introduces SPD-Conv with a scaling factor of 2 for down-sampling. SPD-Conv can retain all the information in the channel dimension during down-sampling, enabling the network to fully utilize the feature information extracted from the backbone network. Both Model 2 and Model 4 are aimed to increase the utilization rate of feature information and improve the detection capability for small objects, thus enhancing the overall performance of the network. Besides, Model 5 builds upon Model 1 by introducing an SimAM mechanism,



(a) Loss curve

**Fig. 9** The curves of training losses and mAPs

which allows the network to focus on important features and suppress irrelevant ones during the object detection process, resulting in a 0.9% increase in mAP. It is worth noting that Model 6 adds the SimAM mechanism after SPD-Conv on the basis of Model 4, and mAP increases from 95.1% to 96.1%, that is because SimAM can accelerate feature fusion after SPD-Conv without increasing network parameters. Based on this, Model 7 adds the feature prediction layer, increasing mAP to 96.7%, and reaching the optimal P-value of 97.7%. Our model, i.e. Model 8, achieves an R of 93.3%, an mAP of 97.1%, and obtains a suboptimal P-value. Considering all the quantitative metrics, the proposed method demonstrates the best performance. In addition, to better observe the entire training process of the ablation experiments, we plot the curves of loss and mAP as shown in Fig. 9.

3.4.2 Comparison with Other Defect Detection Methods

In order to quantitatively evaluate the detection advantages of the improved YOLOv5 network proposed, the detection performance of the proposed algorithm is compared with that of Faster R-CNN, SSD, YOLOv5, YOLOv6 and YOLOv7, which are popular in engineering applications. The comparison results of each network are shown in Table 4 and Fig. 10.

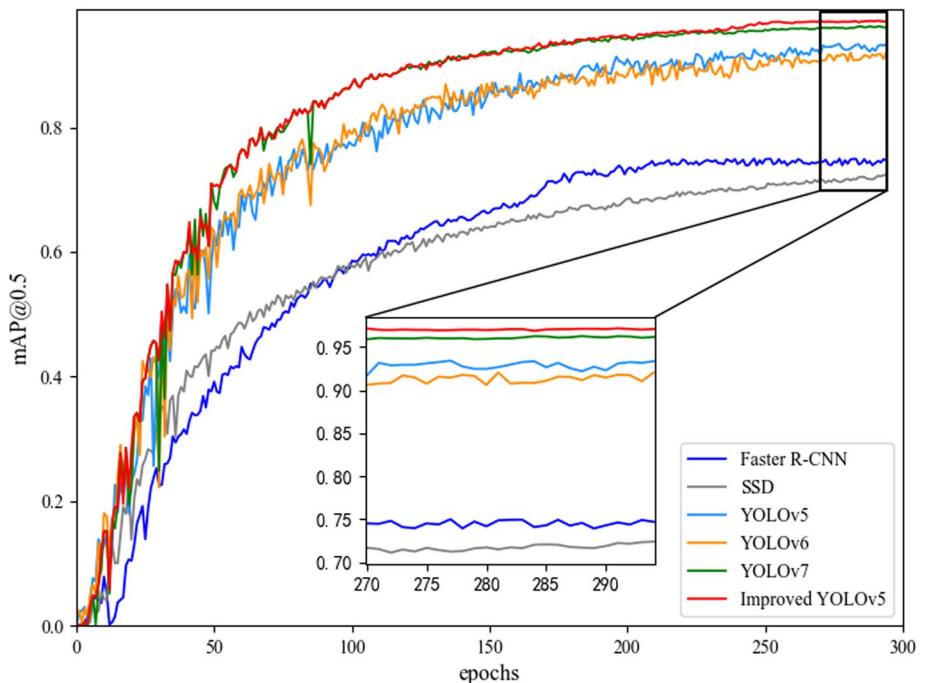
In Table 4, the average accuracy of the Faster R-CNN is 74.8%, which is higher than that of the SSD. Generally speaking, two-stage object detection networks typically achieve higher detection accuracy compared to one-stage networks. This is due to the two-stage approach using a two-part architecture to handle the problem of class imbalance and RPN (Region Proposal Network) to balance positive

Table 4 Comparative experiments

Methods	AP/%							P/%	R/%	mAP@0.5/%
	Nest	Insulator	Break	Defect	Damper	Defective damper				
Faster R-CNN	95.6	87.1	47.9	60.2	73.8	84.3	85.9	73.5	74.8	
SSD	92.3	84.7	46.1	58.0	72.3	79.9	70.6	61.7	72.1	
YOLOv5	99.4	97.2	70.3	99.5	94.9	95.8	92.4	87.9	93.3	
YOLOv6	99.1	96.9	78.3	94.1	95.2	88.4	93.2	88.8	91.9	
YOLOv7	99.5	98.5	87.7	99.0	95.7	96.3	95.5	93.7	96.1	
Proposed method	99.5	98.8	92.6	98.2	96.6	97.1	96.8	93.3	97.1	

Bold font indicates that the model performs the best under the same criterion

Fig. 10 The variation curves of mAPs



and negative samples when fitting bounding boxes. By using a two-stage cascaded approach, these networks are able to more accurately detect objects in complex and cluttered scenes. However, with the iterative optimization of YOLO series detection network, the accuracy of network detection is constantly improved. Among them, the mAP value of YOLOv5 on the NID dataset reaches 93.3%, which represents a significant improvement over Faster R-CNN and SSD. This performance boost is attributed to the integration of a series of optimization techniques, including the application of the focal loss function, which significantly enhances the capability of the network in object detection. Notably, YOLOv5 achieves the highest AP value of 99.5% for insulators with defects among all the algorithms tested. The P and R of YOLOv6 are 93.2% and 88.8% respectively. The average detection accuracy is 91.9%, which is worse than YOLOv5. The mAP scores for YOLOv7 and the proposed method are 96.1% and 97.1%, respectively. In particular, both achieve a detection accuracy of 99.5% for the bird nest category. Moreover, the proposed method demonstrates high detection accuracy across multiple categories, achieving an AP value of 92.6% for small damaged insulators, which represents a significant improvement of 19.6% over that of YOLOv5.

Figure 10 presents the variation curves of the mAPs of the network models during the training process. From the figure, Faster R-CNN converges the slowest and becomes

stable at around 200 epochs. The mAP curve of SSD is below other curves and stays around 72% at the final value. YOLOv5, YOLOv6, YOLOv7, and the improved YOLOv5 converge gradually at about 100 epochs, and the convergence of the improved network is relatively fast, and the average detection accuracy curve finally stays at about 97%. It can be concluded that the improved network has higher parameters and better training advantages.

3.4.3 Qualitative Analysis

In this section, a selection of typical outdoor transmission line images is screened from the testing set to test the proposed method, and compared with Faster R-CNN, SSD, YOLOv5, YOLOv6, and YOLOv7 algorithms. It is worth noting that to better observe the detection results, we enlarges the detected images to different scales.

Figure 11 shows the detection results of the general scene in transmission line inspection. As seen from Fig. 11f, the proposed method recognizes all objects in the image. And except for our algorithm, the defective damper in the lower right position of the image is not detected by other algorithms, because the damper itself has a similar color to the soil and the target is too small. Particularly, in Fig. 11a, b, due to the small color difference between the target and the background, Faster R-CNN and SSD misjudges the intact damper as one with defects.

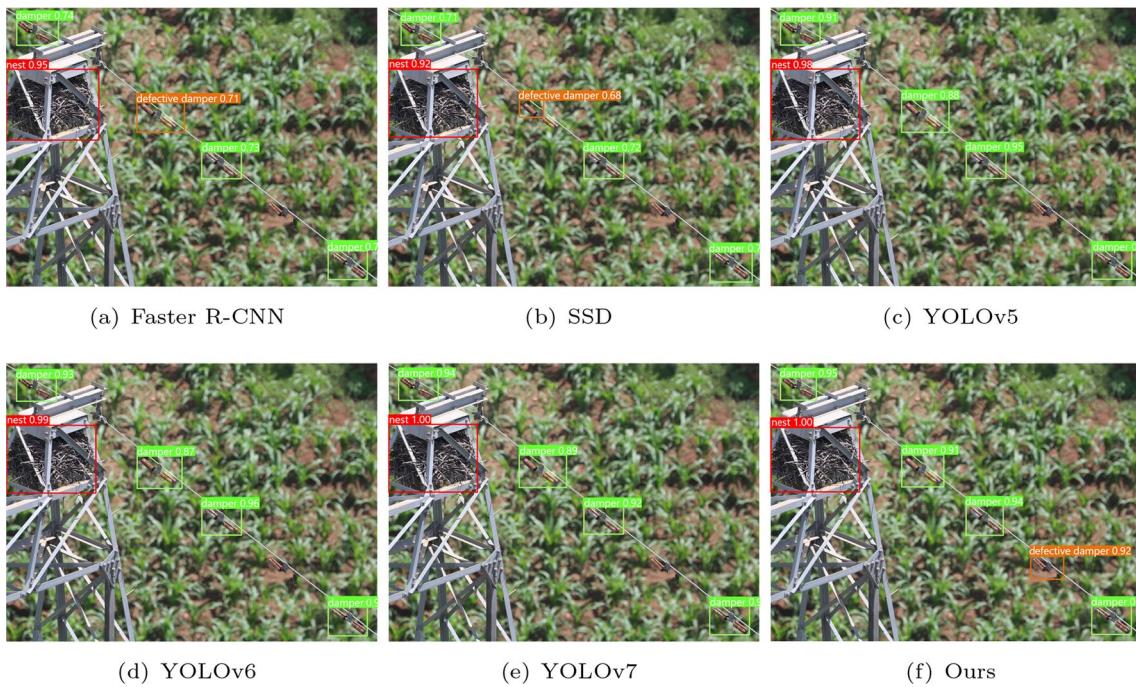


Fig. 11 The detection results under small color difference conditions

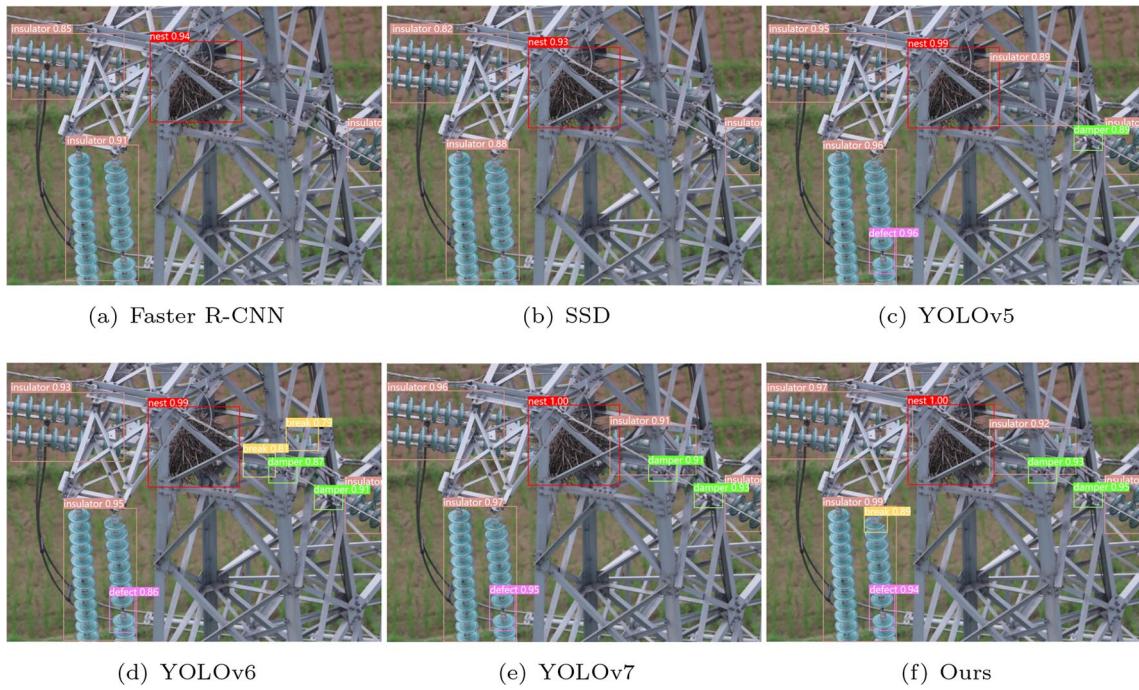


Fig. 12 The detection results under complex and occlusive conditions

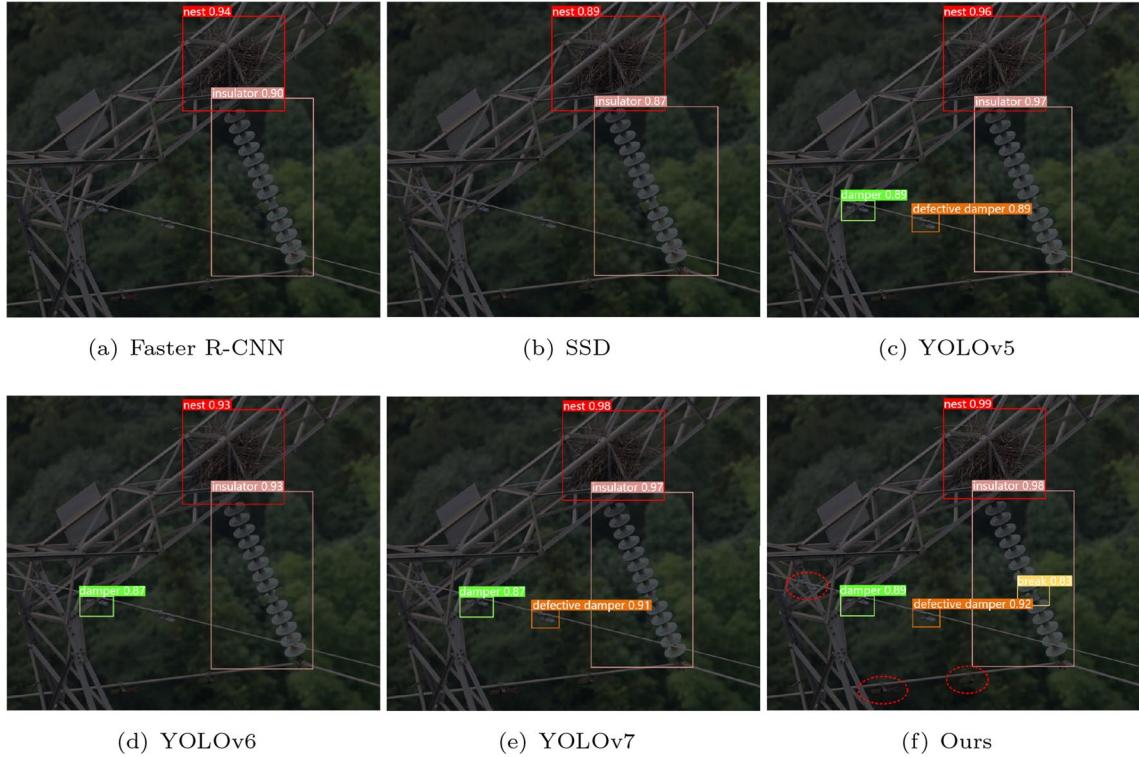


Fig. 13 The detection results under low-lighting conditions

Table 5 Comparison with related works on the CPLID

No	Work	Method	P/%	R/%	mAP/%
1	Tao et al. [28]	ILN	90.4	96.6	93.4
		DDN	91.0	95.8	93.3
2	Li et al. [29]	Faster R-CNN + OHEM	Not given	Not given	90.75
3	Ding et al. [30]	YOLOv5 (EIoU+AFK-MC2)	Not given	Not given	96.0
4	Wang et al. [31]	YOLOv4 + data augmentation	91.0	98.84	99.08
5	The proposed method	Improved YOLOv5	98.6	94.3	97.8

Figure 12 illustrates the detection results of Faster R-CNN, SSD, YOLOv5, YOLOv6, YOLOv7, and the proposed method under a complex detection environment with occlusions. The results indicate that due to occlusion by the high-voltage tower, the insulators on the right side of the image center are not accurately recognized by Faster R-CNN, SSD, and YOLOv6 algorithms. In fact, YOLOv6 incorrectly identifies them as damaged insulators. Furthermore, the small color difference between the two dampers near the right side of the image and the high-voltage tower, as well as the small size of the defective insulator(defect) at the bottom of the image, make it impossible for Faster R-CNN and SSD to recognize them. YOLOv5 also misses one damper in this context. Especially for the damaged insulator(break) on the left side of the image center, all algorithms, including Faster R-CNN, SSD, YOLOv5, YOLOv6, and YOLOv7, fail to accurately locate the detection object owing to its small size. In contrast, as shown in Fig. 12f, the proposed algorithm successfully detects all objects with a high localization accuracy.

What is more, we conduct additional tests to compare the efficacy of the proposed algorithm under low-lighting conditions, the detection results are shown in Fig. 13. Our method can locate the faulty insulator accurately, demonstrating superior detection performance than other algorithms. While YOLOv6, Faster R-CNN, and SSD only recognize limited objects, such as bird nest and insulator, failing to identify the defective damper. Nevertheless, the proposed method also misses several detection objects in Fig. 13f, which are represented by dashed ellipses. This is because the lighting conditions of the test image are poor, resulting in very low contrast between the damper and the background color. Instances of such situations are relatively infrequent during training, resulting in the inability of network to detect them. In addition, more comparative test results are shown in Figs. 14, 15 in the Appendix.

3.4.4 Comparison with Related Works

In this section, to further validate the effectiveness of our algorithm in this paper, experiments will be conducted on the CPLID [28]. The China power line insulator data set (CPLID) is provided by the China Electric Power Research Institute and contains 600 normal insulator images and 248

synthetic defective insulator images. It is a dataset used for insulator defect detection.

Table 5 presents the comparison results between our proposed method and other related works. In this table, four reference results on the CPLID are selected and named NO.1-NO.4, and NO.5 represents the experimental results of our proposed method. The last three columns represent the evaluation metrics of P, R, and mAP values, where “Not given” indicates that the corresponding values are not provided in the original work. The table reveals that Wang et al. [31] achieves the highest mAP value with a score of 99.08%, surpassing the performance of our algorithm by 1.28%. Among them, our method has the highest P value on the CPLID, which is 98.6%.

4 Conclusion

The paper proposes a deep learning detection network based on recursive gating design, which is used for intelligent defect detection of foreign object and power component defect in high voltage transmission lines. Firstly, the key target and defect data sets of high voltage transmission lines are constructed by collecting and collating, which is more suitable for the complex on-site detection environment. Secondly, the K-means algorithm is used to optimize the anchor box of the data set to improve the fitting ability of the network to the detection target. Then, a C3CG module is designed to construct a backbone network based on the correlation of high-order spatial features of recursive gated convolution. Finally, the feature detection scale is added on the original basis, the underlying feature information is fused, and the down-sampling is realized by slice convolution SPD-Conv, and the attention mechanism is added to optimize the detection network. Experimental results show that the detection precision and recall rate of key targets and defects of transmission lines under complex environments are 96.8% and 93.3%, respectively, with mAP of 97.1%. It has high precision and robustness in the outdoor complex detection environment, improves the inspection image detection efficiency and accuracy of transmission lines, and provides a new idea for the application and development of smart grid.

Appendix

The comparative test results between the proposed method and other algorithms.



Fig. 14 The comparative test results

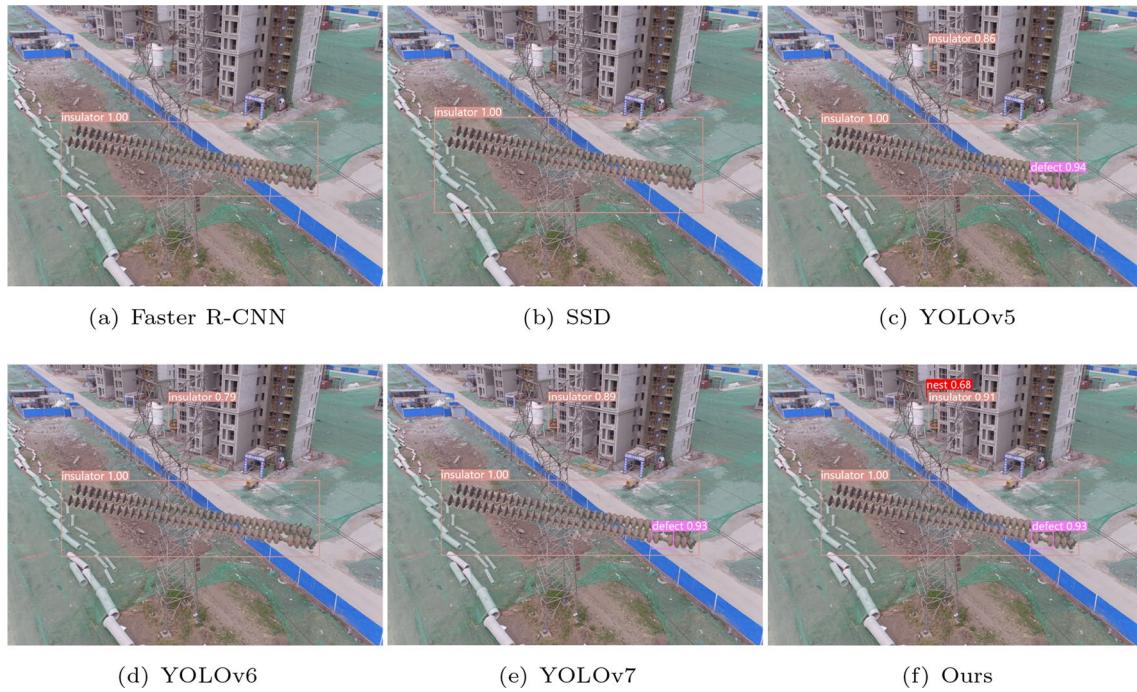


Fig. 15 The comparative test results

Funding This work was in part supported by Key Project of Science and Technology Research Program of Hubei Educational Committee (No. D20211402) and the National Key Research and Development Program (No. 2022YFE0209900).

Declarations

Conflict of interest The authors declare no competing interests.

References

- Zhao Z, Qi H, Qi Y, Zhang K, Zhai Y, Zhao W (2020) Detection method based on automatic visual shape clustering for pin-missing defect in transmission lines. *IEEE Trans Instrum Meas* 69(9):6080–6091
- Ma Y, Li Q, Chu L, Zhou Y, Xu C (2021) Real-time detection and spatial localization of insulators for uav inspection based on binocular stereo vision. *Remote Sensing* 13(2):230
- Mohamed N, Al-Jaroodi J, Jawhar I, Idries A, Mohammed F (2020) Unmanned aerial vehicles applications in future smart cities. *Technol Forecast Soc Chang* 153:119293
- Liang H, Zuo C, Wei W (2020) Detection and evaluation method of transmission line defects based on deep learning. *IEEE Access* 8:38448–38458
- Zhang Y-C, Liang Z, Tan M (2004) Mobile robot for overhead powerline inspection: a review. *Robot* 26(5):467–473
- Zhang Y, Yuan X, Fang Y, Chen S (2017) Uav low altitude photogrammetry for power line inspection. *ISPRS Int J Geo Inf* 6(1):14
- Li Y, Dai L, Ni H, Kong C, Chen X (2023) Focus on local: transmission line defect detection via feature refinement. *SIViP* 17(1):31–37
- Zaidi SSA, Ansari MS, Aslam A, Kanwal N, Asghar M, Lee B (2022) A survey of modern deep learning based object detection models. *Digital Signal Processing* 103514
- Tulbure A-A, Tulbure A-A, Dulf E-H (2022) A review on modern defect detection models using dcnn-s-deep convolutional neural networks. *J Adv Res* 35:33–48
- Wei Z-X (2022) Composite insulator defect identification and quantitative method based on random hough transform ellipse detection. *J Phys: Conf Ser*, vol. 2170, p 012021. IOP Publishing
- Tan P, Li X-F, Xu J-M, Ma J-E, Wang F-J, Ding J, Fang Y-T, Ning Y (2020) Catenary insulator defect detection based on contour features and gray similarity matching. *J Zhejiang Univ-Sci A* 21(1):64–73
- Ren S, He K, Girshick R, Sun J (2015) Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv Neural Inf Process Syst* 28
- Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 779–788
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, Berg AC (2016) Ssd: single shot multibox detector. In: Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer, pp 21–37
- Zhao W, Xu M, Cheng X, Zhao Z (2021) An insulator in transmission lines recognition and fault detection model based on improved faster r-cnn. *IEEE Trans Instrum Meas* 70:1–8
- Qiu Z, Zhu X, Liao C, Shi D, Qu W (2022) Detection of transmission line insulator defects based on an improved lightweight yolov4 model. *Appl Sci* 12(3):1207
- Liu Z, Wu G, He W, Fan F, Ye X (2022) Key target and defect detection of high-voltage power transmission lines with deep learning. *Int J Electr Power Energy Syst* 142:108277
- Hao K, Chen G, Zhao L, Li Z, Liu Y, Wang C (2022) An insulator defect detection model in aerial images based on multiscale feature pyramid network. *IEEE Trans Instrum Meas* 71:1–12
- Zhang H, Wu C, Zhang Z, Zhu Y, Lin H, Zhang Z, Sun Y, He T, Mueller J, Manmatha R (2022) et al: Resnest: Split-attention networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 2736–2746
- Li Y, Ni M, Lu Y (2022) Insulator defect detection for power grid based on light correction enhancement and yolov5 model. *Energy Rep* 8:807–814
- Rao Y, Zhao W, Tang Y, Zhou J, Lim S-N, Lu J (2022) Hornet: efficient high-order spatial interactions with recursive gated convolutions. *arXiv preprint arXiv:2207.14284*
- Wang C-Y, Liao H-YM, Wu Y-H, Chen P-Y, Hsieh J-W, Yeh I-H (2020) Cspnet: a new backbone that can enhance learning capability of cnn. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, pp 390–391
- Woo S, Park J, Lee J-Y, Kweon IS (2018) Cbam: convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV), pp 3–19
- Yang L, Zhang R-Y, Li L, Xie X (2021) Simam: a simple, parameter-free attention module for convolutional neural networks. In: International conference on machine learning, pp 11863–11874
- Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S (2017) Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2117–2125
- Li H, Xiong P, An J, Wang L (2018) Pyramid attention network for semantic segmentation. *arXiv preprint arXiv:1805.10180*
- Sunkara R, Luo T (2022) No more strided convolutions or pooling: a new cnn building block for low-resolution images and small objects. *arXiv preprint arXiv:2208.03641*
- Tao X, Zhang D, Wang Z, Liu X, Zhang H, Xu D (2018) Detection of power line insulator defects using aerial images analyzed with convolutional neural networks. *IEEE Trans Syst Man Cybern: Syst* 50(4):1486–1498
- Li X, Su H, Liu G (2020) Insulator defect recognition based on global detection and local segmentation. *IEEE Access* 8:59934–59946
- Ding J, Cao H, Ding X, An C (2022) High accuracy real-time insulator string defect detection method based on improved yolov5. *Front Energy Res* 10:928164
- Wang S, Zou X, Zhu W, Zeng L (2023) Insulator defects detection for aerial photography of the power grid using you only look once algorithm. *J Electr Eng Technol*, 1–14

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Shanshan Wang received the B.S. degree from the Huazhong University of Science and Technology, Wuhan, China, and the M.S. degree as well as the Ph.D. degree from the Huazhong University of Science and Technology, Wuhan, China. She is currently an associate professor at Hubei University of Technology. Her research interests include machine learning and deep learning, medical image analysis, ultrasound imaging, image processing, etc.

Weiwei Tan is currently studying for a master's degree at the school of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan, Hubei Province, China. The main research interests are deep learning, image processing, and object detection.

Tengfei Yang is currently pursuing the master's degree in Electrical Engineering at Hubei University of Technology, China since 2021. His research interests include deep learning, image processing, and object detection.

Liang Zeng received the B.S. degree and the M.E. degree from Wuhan University of Technology, Wuhan, China. his Ph.D. degree from Huazhong University of Science and Technology, Wuhan, China. He is currently a professor at Hubei University of Technology, Wuhan. His research interests include artificial intelligence, evolutionary algorithms and their applications, modeling and optimization of complex systems, etc.

Wenguang Hou received the B.S. degree from the Wuhan University, Wuhan, China, in 1997, and the M.S. degree as well as the Ph.D. degree from the Wuhan University, Wuhan, China. He is currently an Associate Professor and Doctor Tutor of college of life science and technology, Huazhong University of Science and Technology. His research interests include machine learning and deep learning, medical image analysis, ultrasound imaging, image processing, etc.

Quan Zhou received Ph.D. degree at Huazhong University of Science and Technology, Wuhan, China. He is currently a post doctor at Huazhong University of Science and Technology, Wuhan, China. His research focuses on deep learning and medical image processing.