

## Research Article

# Fast Detection of Defective Insulator Based on Improved YOLOv5s

Zhao Liqun <sup>1</sup>, Zou Mengjun <sup>1</sup>, Cui Ying <sup>2</sup>, and Jia Yanfei <sup>3</sup>

<sup>1</sup>Key Laboratory of Modern Power System Simulation and Control & Renewable Energy Technology, Ministry of Education (Northeast Electric Power University), Jilin 132013, China

<sup>2</sup>Guangdong Electric Power Corporation, Zhuhai Power Supply Bureau, Zhuhai 519000, China

<sup>3</sup>College of Electrical and Information Engineering, Beihua University, Jilin 132013, China

Correspondence should be addressed to Zhao Liqun; zhaoliqun@163.com

Received 29 July 2022; Accepted 16 August 2022; Published 3 September 2022

Academic Editor: Nian Zhang

Copyright © 2022 Zhao Liqun et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Defective insulator detection is an essential part of transmission line inspections based on unmanned aerial vehicles. It can timely discover insulator defects and repair them to avoid a power transmission accident. The detection speed of defective insulators based on artificial intelligence directly affects inspection efficiency. To improve the detection speed of defective insulators based on YOLOv5s, an improved detection method with faster detection speed and acceptable precision is proposed. First, a new ResNet unit with three branches is designed based on depthwise separable convolution with kernel three and average pooling. To reduce parameters, the new ResNet unit is used to replace the original ResNet unit used in the CSP1\_X module in YOLOv5s. Besides, we also introduce channel shuffle in the CSP1\_X module to facilitate the flow of feature information from different channels. Second, a new residual CBL module is designed based on depthwise separable and standard convolution. The new residual CBL module is used to replace the two CBL modules used in the CSP2\_X module in YOLOv5s to reduce parameters and extract more useful features. Third, we design a separate, coordinated attention module by introducing location information into channel attention. The new attention module is added to the end of the CSP2\_X module to improve the ability to extract insulator location information. Besides, we also use convolution to replace the focus model to reduce computation. Compared with defective insulator detection methods, the proposed method has smaller parameters, floating-point operations per second, and higher frames per second. Although it has lower mean precision, it has a faster detection speed. Besides, the increase in detection speed is greater than the decrease in mean precision.

## 1. Introduction

Unmanned aerial vehicle has been widely used in transmission line inspection to improve inspection efficiency and reduce the workload for inspectors [1, 2]. In the traditional transmission line inspection based on an unmanned aerial vehicle, the inspector detects the defective transmission devices by watching the screen. It takes a lot of time to detect faulty devices, which is affected by the screen size and the light intensity. With the development of artificial intelligence technology, the deep learning method has been widely used in object tracking [3, 4], image super-resolution reconstruction [5], image dehazing [6], and defective transmission device detection [7]. The images of transmission line devices captured by the unmanned aerial vehicle are transmitted via a wireless mobile communication network to

a server with high computing power. The captured images are detected by artificial intelligence methods deployed on the server. It consumes much time to transmit images to serves, affecting transmission devices' detection speed. It cannot be used in detection areas without wireless mobile network coverage.

With the development of portable edge computing devices, the captured images can be detected by artificial intelligence methods deployed on the mobile edge computing device [8]. The captured images are transmitted to portable edge computing devices carried by inspectors with short-range wireless communication. It is not dependent on the mobile network and has a greater range of applications. The computation power of portable edge computing devices is limited, which affects the detection speed. In order to improve detection speed, one is to improve the computation

power of portable edge computing devices, and the other is to reduce the computation of the artificial intelligence methods. The cost of reducing computation is lower. Therefore, we mainly focus on how to reduce computation.

Object detection based on deep learning can be divided into two categories: one-stage detection method and two-stage detection method. In the one-stage method, it directly generates detection boxes and classifies the objects without generating region proposals. The two-stage method divides the detection process into two stages. It generates region proposals in the first stage, regresses the bounding box and candidate regions, and classifies the objects in the second stage. The one-stage method focus on improving detection speed. The two-stage method focus on improving detection precision. The two-stage method is mainly developed on the servers with high computing power. It is more suitable for offline object detection. The one-stage method has lower computation than two-stage method. It is more suitable for online object detection. Although it has been widely used in defective insulator detection, the detection is still required to be improved to meet the practical requirements.

YOLOv5 is one of the one-stage methods [9]. Although it is named YOLOv5, it is designed based on YOLOv3 [10] and is unrelated to YOLOv4 [11]. The YOLOv5 has four versions that are YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. The YOLOv5s has a smaller computation and lower precision. The YOLOv5x has a larger computation and higher precision. The YOLOv5s is more suitable for deploying on the portable edge computing device. To improve the defective insulator detection speed of YOLOv5s with acceptable precision, we propose an improved YOLOv5s.

For this paper, the main contributions are as follows:

- (1) To reduce computation and improve the detection speed of defective insulators, we design a new ResNet unit with two branches, and use it to replace the original ResNet unit used in CSP1\_X module in YOLOv5s. The new ResNet is based on depthwise separable convolution with kernel three and average pooling. It has smaller computation than the original ResNet unit In YOLOv5s. Besides, to extract more useful features from different channels, we also introduce the channel shuffle operation into CSP1\_X module in YOLOv5s.
- (2) To further reduce the computation, we also design a residual CBL module and use it instead of two original CBL modules in CSP2\_X module in YOLOv5s. The residual CBL module has smaller computation than the original two CBL modules. Besides, to extract more useful features from different channels, we also introduce the channel shuffle operation into CSP2\_X module in YOLOv5s to facilitate the flow of feature information from different channels.
- (3) To reduce background interference and make the network pay more attention to useful information, we design a separate embedded coordinated attention module and introduce it into the CSP2\_X module to extract more useful features. Besides, to

reduce computation, we also use convolution with kernel 3 to replace the focus.

The traditional transmission line inspection based on unmanned aerial vehicle requires a human eye view of the defective insulators. To improve the inspection efficiency based on unmanned aerial vehicle, artificial intelligence methods are deployed on servers or ground service stations to automatically detect defective insulators from the captured image by unmanned aerial vehicles. Although many artificial intelligence methods have higher precision, they have slower detection speed that directly affects inspection efficiency. To improve defective insulator detection speed with acceptable detection precision, we propose an improved YOLOv5s with lower computation for defective insulator detection to improve the inspection efficiency. The structure of the rest of this article is organized as follows. Related work is provided in Section 2. Our proposed method are introduced in Section 3. The experimental results and discussions are reported in Section 4. Finally, conclusions are given in Section 5.

## 2. Related Work

Object detection method based on deep learning has been widely used in defective insulator and transmission device detection. There are two categories of object detection models. The first category is two-stage detection models, such as Faster R-CNN [12], Mask R-CNN [13], Cascade R-CNN [14], and Sparse R-CNN [15]. The second category is one-stage detection models, such as the SSD series method [16, 17] and the YOLO series method [9–11]. The two-stage has a larger computation and is more suitable for offline detection. The one-stage has smaller computation and is more suitable for online detection. Therefore, the defective insulator detection method based on a one-stage method is more suitable for developing a portable edge computing device with limited computation power.

The YOLO series method is the most representative method of the one-stage detection method. The YOLOv1-method was first proposed by Redmon et al. [18]. It creatively combined candidate area and object recognition into one stage. It transformed the object recognition problem into a regression problem and directly predicted the location and class of the object using a depth convolution neural network. Compared with the two-stage method, it has lower computation and acceptable accuracy. Due to these advantages, the YOLO method attracts much attention from many scholars and has become an important branch of object detection research based on deep convolution neural networks. To improve the detection speed and precision of YOLOv1, Redmon et al. [19] proposed YOLOv2. In YOLOv2, batch normalization is introduced into the convolution layer. The new convolution layer consists of convolution, batch normalization, and LeakyReLU activating function. In the VOC2007 dataset, the mean average precision is improved from 63.4% to 65.8%. In order to solve the problem caused by the different training and detection image sizes, YOLOv2 fine-tuned the classification network that had been trained on  $244 \times 244$  low-resolution images on

448 × 448 high-resolution images. After fine-tuning, the final global average pooling and softmax layers are removed as the final backbone network. Then mean average precision is improved from 65.8% to 69.5%. It also introduced anchor boxes inspired by Faster R-CNN to improve precision. The size of the output feature map is 13 × 13. Each cell contains five anchor boxes to predict five bound boxes. Besides, YOLOv2 designed a new network named Darknet-19 and used it as a backbone. In 2018, Redmon proposed YOLOv3, which is also the last version proposed by Redmon [10]. In order to extract more useful features, YOLOv3 used the Darknet-53 as the backbone and introduced feature pyramid networks to fuse more features. The number of network layers increased and contained many residual networks in Darknet-53. Besides, it also proposed binary cross-entropy loss for classification. Although Redmon withdraws from research in artificial intelligence, many improved YOLOv3 are proposed by scholars.

Yang et al. proposed GC-YOLOv3 based on YOLOv3 to improve the mean average precision [20]. They designed a cascading network that consisted of learnable semantic fusion and used a global self-attention mechanism to extract more useful information. Qu et al. proposed an improved YOLOv3 with auxiliary networks for remote sensing image detection to improve detection precision and detection speed [21]. They used an image blocking module to feed fixed images and DIOU to replace IOU in YOLOv3. Besides, they used a convolutional block attention module to connect the backbone network and designed an auxiliary network. In order to improve detection speed, the adaptive feature fusion method was also introduced into the improved YOLOv3. In order to improve detection speed, Yin et al. proposed Faster-YOLO [22]. They used ELELEM-AE joint network and DRKCELM network to design a feature extractor. The feature extractor integrates the advantages of ELM-EA and ELM-LRF. The detection speed of Faster-YOLO is two times faster than YOLOv3. In order to improve detection precision, Cai et al. proposed a modified YOLOv3 [23] based on MobileNetV1. They used the MobileNetV1 to replace the backbone of YOLOv3 and optimized the feature map size according to the detection results.

In 2020, Alexey et al. [11] proposed the YOLOv4 method base on YOLOv3. They used the cross-stage partial connections network to replace the residual block in YOLOv3 to design the backbone of YOLOv4. The path aggregation network also was used to fuse more features. They used spatial pyramid pooling to realize multi-scale features fusion and mish function to replace the LeakyReLU function as a new activate function. Besides, they used the mosaic data argument method to improve detection precision. The YOLOv4 has a faster detection speed and precision. To improve the detection speed of YOLOv4, Deng et al. [24] proposed an improved YOLOv4. They used feature pyramid networks and atrous spatial pyramid pooling to modify the MobileNetV3 to improve real-time and feature extraction ability. The improved MobileNetV3 is used as the backbone of YOLOv4 to reduce computation. They also introduced a convolutional block attention module to YOLOv4 to extract more useful features. The original team of YOLOv4

proposed scaled-YOLOv4 based on YOLOv4-CSP to make the model can be developed on different devices [25]. To further improve the performance of YOLOv4, some improved methods based on YOLOv4 have also been proposed [26–28]. To improve the one-stage method performance, Glenn proposed YOLOv5 based on YOLOv3 [10]. YOLOv5 designed two types of cross-stage partial (CSP) Networks that are CSP1\_X and CSP2\_X. The CSP1\_X modules were used in the backbone, and the CSP2\_X were used in the Neck part of YOLOv5. YOLOv5 has four different version networks that are YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. The different version of YOLOv5 has different depth CSP modules. YOLOv5 used the LeakyReLU function to activate the function in middle/hidden layers and the sigmoid function as an activated function in detection layers. Besides, YOLOv5 used GIOU loss as the loss function of the bounding box. YOLOv5s has a faster detection than YOLOv4 and YOLOv3. It is more suitable for developing portable edge computing devices.

Due to the advantage of the YOLO series method, they have been widely used in transmission devices and defective insulator detection. Liu et al. [29] proposed an aerial insulator image detection method based on YOLOv3 for high-voltage transmission lines. A cross-stage partially densely connected module was proposed to solve the feature reuse and propagation of feature layers in low-resolution images. It had higher detection precision defective insulator detection in complex transmission line backgrounds than YOLOv3 and YOLOv4. Qiu et al. [30] proposed a defective insulator detection algorithm based on YOLOv4. They used the Graph Cut data enhancement method to produce a new dataset, and the Laplace sharpening method was used to preprocess the insulator dataset images. To make the algorithm more lightweight, they used MobileNet as the backbone network of YOLOv4. It had a faster detection speed than YOLOv4. He et al. [31] proposed a self-exploding insulator detection algorithm based on YOLOv4. They proposed a new feature fusion structure and an improved SE attention mechanism, which effectively suppressed useless features and achieved higher detection accuracy than YOLOv4. Feng et al. [32] verified in detail the performance of YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x on the public dataset China Power Line Insulation Dataset (CPLID), and the experimental results showed that YOLOv5 performed well on this dataset, especially YOLOv5x, with a detection accuracy of 95.5%, which can effectively identify defective insulators. Since the YOLOv5x model has a large number of parameters, Lan et al. [33] selected YOLOv5s with a smaller number of parameters as the baseline model, making it lighter by replacing the original CSP structure with the Ghost module and adding CBAM attention to improve the detection accuracy. To further improve the detection accuracy of defective insulators, some other methods based on YOLO have also been proposed [34–36].

### 3. Improved YOLOv5s

*3.1. Improved CSP Module.* Two types of CSP modules are given as CSP1\_X and CSP2\_X in YOLOv5s. The CSP1\_X

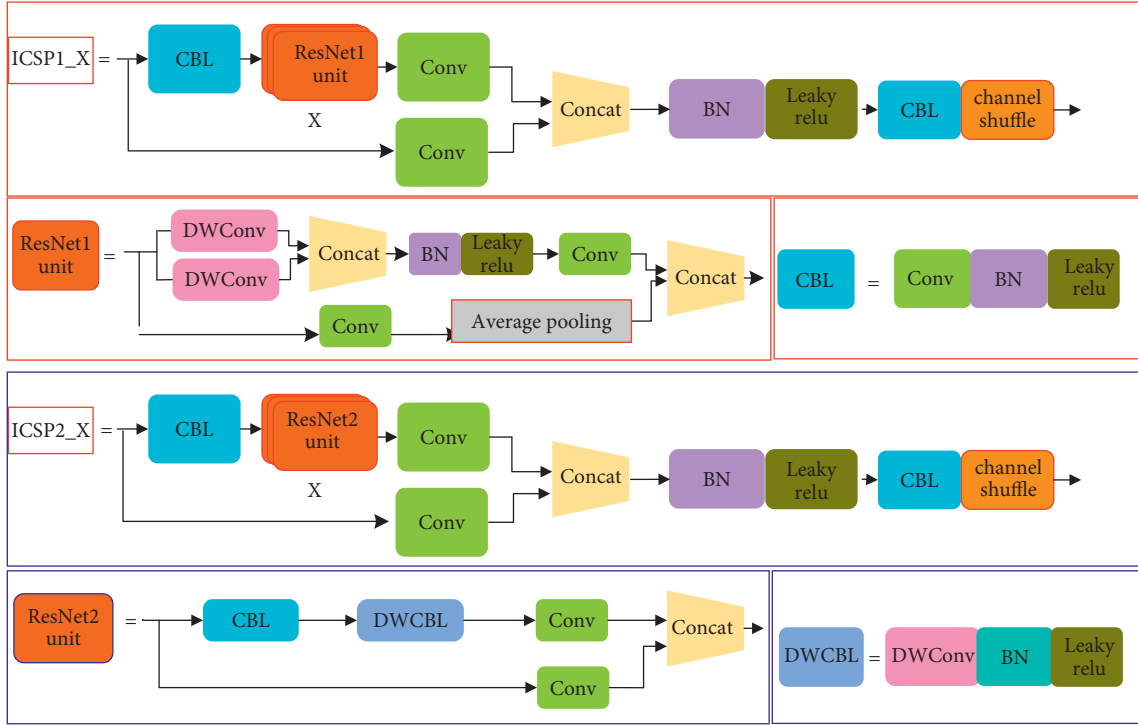


FIGURE 1: Improved CSP1\_X and CSP2\_X.

module is used in the backbone to extract features and the CSP2\_X module is used in the Neck to fuse features. The CSP1\_X consists of two branches. The first branch mainly consists of CBL module, ResNet units, and a convolution layer. The second branch only consists of a convolution layer. The output features of two branches are fused by concatenate operation. The ResNet unit computation directly affects detection speed.

To reduce the computation of CPS1\_X, we designed a new ResNet unit and replaced the original ResNet unit in CSP1\_X. The proposed ResNet unit is named as ResNet1 unit and is shown in Figure 1. The proposed ResNet1 unit consists of two branches that we name the upper branch and lower branch, respectively. The upper branch contains two parallel depthwise separable convolutions (DWConv). The kernel sizes of the two depthwise separable convolutions are 3, and the number of output channels is half the number of input channels. We use the two parallel depthwise separable convolutions to extract different features. Therefore, we use concatenate operation to connect the output feature maps of two DWConv to keep the number of channels constant. To avoid gradient explosion or gradient disappearance and improve training speed, we add batch normalization (BN) and LeakyReLU activating function in the upper branch. Besides, we use a  $1 \times 1$  convolution to reduce the number of output channels to halve the original number. The lower branches contain a  $1 \times 1$  convolution and an average pooling. The  $1 \times 1$  convolution is used to reduce the number of channels to make the lower branch have the same number of channels as the upper branch.

The average pooling is used to reduce the value of the background feature, which is useful for separating the

defective insulator and background information. In the end, the upper and lower branches are connected by concatenating operation to construct the complete ResNet1 unit. We use the proposed ResNet1 unit to replace the original ResNet unit to reduce the computation of CSP1\_X. In the end, we use the BN module, LeakyReLU activating function, and CBL module to extract features from the concatenated feature map. The CBL module consists of a  $1 \times 1$  standard convolution, a BN module, and a LeakyReLU activating function. They are used to increase network depth to make the network extract more detailed information about the defective insulator. The improved CPS1\_X is named ICSP1\_X which is shown in Figure 1. The X is the number of ResNet1 unit used in ICSP1\_X. The X is different in different modules. In YOLOv5s, two different ICSP1\_X modules are present, i.e., ICSP1\_1 and ICSP1\_3. The numbers of ResNet1 unit of ICSP1\_1 and ICSP1\_3 is 1 and 3, respectively. The ICSP1\_X module does not change the number of feature map channels. They are only used in the backbone of the network to extract features.

To further reduce the computation of the network, we also design a new residual module to replace the two CBL modules in CSP2\_X. We name the proposed residual unit and improved CSP2\_X as ResNet2 unit and ICSP2\_X, respectively. The ResNet2 unit and ICSP2\_X are shown in Figure 1. The proposed ResNet2 unit also contains two branches. The upper branch includes a CBL module, a DWCBL module, and a  $1 \times 1$  convolution. The CBL module consists of a  $1 \times 1$  standard convolution, a BN module, and a LeakyReLU activating function. The CBL module is used to increase the depth of the network, which is useful for extracting more detailed information about the defective

insulator. The number of output channels for the CBL module is half the number of input channels in the ResNet2 unit. The DWCBL module consists of  $3 \times 3$  depthwise separable convolutions, a BN module and a LeakyReLU activating function. The input feature map of DWCBL module has the same number of channels as the output feature map of DWCBL module. The last  $1 \times 1$  standard convolution in ResNet2 unit is also used to increase the network depth. The lower branch only consists of a  $1 \times 1$  standard convolution. The number of output channels for the  $1 \times 1$  standard convolution is half the number of input channels. In the end, the upper and lower branches are connected by concatenating operation to construct the complete ResNet2 unit. The input feature map of the proposed ResNet2 unit module has the same number of channels as the output feature map of the ResNet2 unit. We use the proposed ResNet2 unit to replace two CBL modules in the original CSP2\_X to obtain the ICSP2\_X module. The ICSP2\_X also consists two branches that are named upper branch and lower branch, respectively. The upper branch consists of a CBL module,  $X$  ResNet2 unit and a  $1 \times 1$  standard convolution. The number of output channels for the upper branch is also the is half the number of input channels. The lower branch of ICSP2\_X is only a  $1 \times 1$  standard convolution to reduce the number of input channels. The concatenating operation is also used to connect upper and lower branches. The same BN module, LeakyReLU activating function, and CBL module are also used to extract features from the concatenated feature map in the ICSP2\_X module. Besides, we also introduce the channel shuffle into ICSP2\_X to facilitate the flow of feature information from different channels.

In the ICSP1\_X and ICSP2\_X, we use depthwise separable convolution to replace the standard  $3 \times 3$  convolution. The FLOPs of standard convolution can be expressed as:

$$N_c = 2 \times K^2 \times C_{in} \times C_{out} \times H \times W, \quad (1)$$

where  $K$  is the kernel size of standard convolution,  $C_{in}$  and  $C_{out}$  are the number of channels for input and output, respectively, and  $H$  and  $W$  are the width and height of the feature map. The FLOPs of depthwise separable convolution can be expressed as:

$$D_c = 2 \times (K^2 \times C_{in} + C_{out} \times C_{in}) \times H \times W. \quad (2)$$

The FLOPs ratio of standard convolution and depthwise separable convolution is:

$$\begin{aligned} \frac{D_c}{N_c} &= \frac{2 \times (K^2 \times C_{in} + 1^2 \times C_{in} \times C_{out}) \times H \times W}{2 \times K^2 \times C_{in} \times C_{out} \times H \times W} \\ &= \frac{K^2 + C_{out}}{K^2 * C_{out}}. \end{aligned} \quad (3)$$

In YOLO5s, the numbers of output channels are 128, 256, and 512 for different modules. The kernel sizes are 1 or 3 for other convolutions. The number of output channels is much larger than kernel size, so the (3) can be simplified as:

$$\frac{D_c}{N_c} = \frac{K^2 + C_{out}}{K^2 * C_{out}} \approx \frac{C_{out}}{K^2 * C_{out}} = \frac{1}{K^2}. \quad (4)$$

The original ResNet unit of CSP1\_X contains a  $1 \times 1$  convolution and a  $3 \times 3$  convolution. The proposed ResNet unit of ICSP1\_X contains two  $1 \times 1$  convolutions and two  $3 \times 3$  depthwise separable convolutions. For the CSP1\_X, the computation of the proposed ResNet1 unit and original ResNet unit is about  $2/9$ . For the CSP2\_X, the computation of the proposed ResNet2 unit and CBL model is about  $1/9$ . Our proposed two networks that are the ResNet1 unit and ResNet2 unit reduce the computation of CSP1\_X and CSP2\_X. The proposed ICSP1\_X and ICSP2\_X have smaller computations than the original CSP1\_X and CSP2\_X.

### 3.2. Separate Embedded Coordinated Attention Module.

The backbone of our improved YOLOv5s based on our proposed ICSP1\_X and ICSP2\_X has few parameters and computation but also reduces feature extraction ability. Inspired by [37], to balance the detection speed and precision, we designed a separated embedded coordinated attention module named SCA module and introduce it into ICSP1\_X module and ICSP2\_X module to extract more useful defective insulator features. The proposed SCA module is shown in Figure 2. It consists of two branches: the upper branch consists of short cut to reserve input feature map. The lower branch is used to compute the weight of useful features; the lower branch has two average pooling operations: adaptive average pooling (H) and adaptive average pooling (W). The kernel sizes of adaptive average pooling (H) and adaptive average pooling (W) are  $H \times 1$  and  $1 \times W$ , respectively. They encode each channel along the horizontal and vertical coordinates, respectively. We can obtain two separate position-aware feature maps from the two adaptive average pooling operations. Therefore, we can capture long-range dependencies of feature information along one spatial direction while retaining precise location information in the other spatial direction. To make the network have better expression ability. We use CBH modules consisting of a convolution layer, a BN layer and a  $H\_sigmoid$  activate function. Without changing the size of the direction-aware feature map, we use the  $1 \times 1$  convolution layer to increase the depth of the network, use BN layer to improve the generalization ability of the network and use the  $H\_sigmoid$  active function to increase the nonlinear representation ability of the network. Then two  $1 \times 1$  convolutions are used to reduce the number of channels to half the original number. The output feature of two branches that contain two adaptive average pooling are connected by concatenate operation. The CBH module is also used to fuse the feature obtained by concatenate operation. The split method is used to separate the feature map from the horizontal coordinate and vertical coordinate. In the end, we can obtain the weight of the horizontal coordinate, and the weight of the vertical coordinate of the feature map by the CBH module and sigmoid activate function. We multiply the input feature map by two weights to get the output feature map.

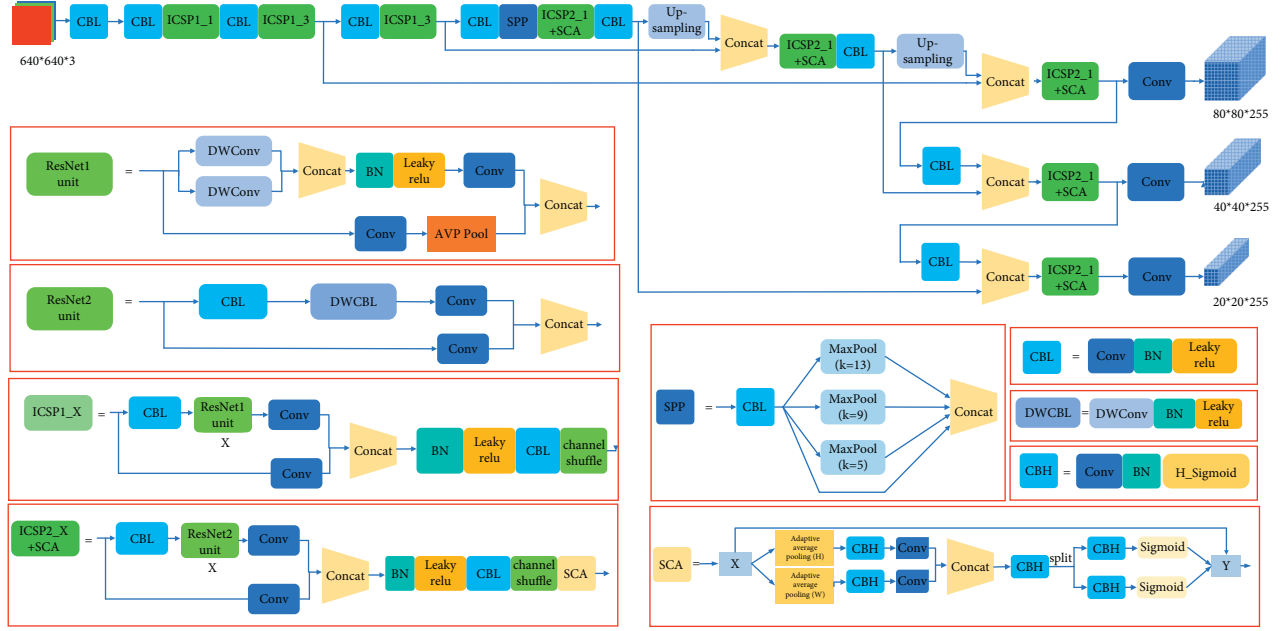


FIGURE 2: Improved YOLOv5s.

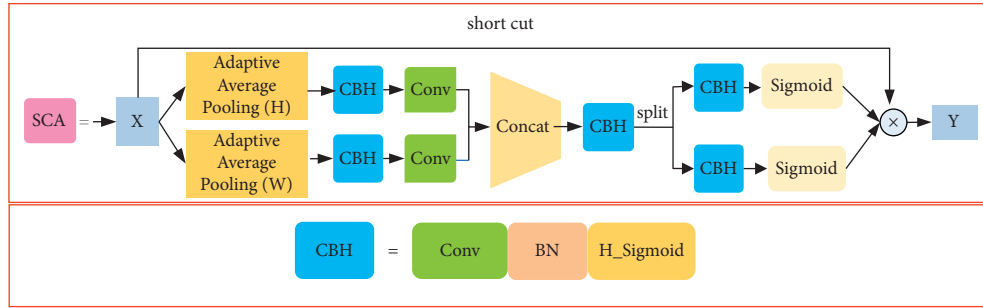


FIGURE 3: Proposed separate embedded coordinated attention module.

To better explain the model, we suppose that the input feature map  $X \in R^{C \times H \times W}$ . The number of channels is  $C$ , and the size of the feature map is  $H \times W$ . Output feature map of adaptive average pooling (H) and adaptive average pooling (W) in the  $c$ th channel can be expressed as  $Z_c(H)$  and  $Z_c(W)$ . They can be expressed as followings:

$$\begin{aligned} Z_c(H) &= \frac{1}{W} \sum_{0 \leq i < W} X_c(H, i) \\ Z_c(W) &= \frac{1}{H} \sum_{0 \leq i < H} X_c(i, W), \end{aligned} \quad (5)$$

where  $X_c$  is the input feature map in the  $c$ th channel. The sizes of  $Z_c(H)$  and  $Z_c(W)$  are  $H \times 1$  and  $1 \times W$ , respectively. The  $1 \times 1$  convolution is used to halve the number of output channels of CBH in Figure 3. The concat operation connects the two output feature maps obtained from CBH and a  $1 \times 1$  convolution. It can be expressed as follows.

$$X_1 = [f^{1 \times 1}(CBH(Z(H))), f^{1 \times 1}(CBH(Z(W)))], \quad (6)$$

where  $f^{1 \times 1}$  is the  $1 \times 1$  convolution,  $[\ ]$  is the concatenating operation.  $X_1$  is the output feature map of concat operation in Figure 2. Next, we use the split function to separate the output feature of the CBH module that is connected with concat operation module into two parts. In the end, we obtain the weights of two parts. The sizes of weight are  $H \times 1$  and  $1 \times W$ , respectively. The size of the final weight obtained by multiplying two weight vectors is  $H \times W$ . Finally, we multiply the input feature map  $X$  with the final weight to get the output feature map.

**3.3. Improved Focus Module.** In YOLOv5s, it uses focus to realize down-sampling without losing information. In focus, it firstly uses a slicing operation to expand the input channel four times. Second, it uses a  $3 \times 3$  convolution to obtain a down-sampling feature map. There is a large computation for the slicing operation. To reduce computation, we use a  $3 \times 3$  convolution with a down-sampling function to replace the focus module. We suppose the input image size is  $640 \times 640 \times 3$  and the size of the output feature map size is

TABLE 1: Pseudo code of our proposed method.

---

```

 $\theta \leftarrow$  Improved YOLOv5s Training ( $z_1: N_{data}, x_1: N_{data}, \theta$ )
/* Training network model
Input:  $\{(z_n, x_n)\}_{n=1}^{N_{data}}$ , a dataset of sequence pairs
Input:  $\theta$ , initial model parameters
Output:  $\theta$ , the trained parameters
Hyperparameters:  $N_{epochs} \in \mathbb{N}$ 
For  $i = 1, 2, \dots, N_{epochs}$  do
For  $n = 1, 2, \dots, N_{data}$  do
 $\theta \leftarrow$  Improved YOLOv5s ( $(z_n, x_n, \theta)$ )
 $loss(\theta) \leftarrow l_{con}(\theta) + 0.5 \times l_{cls}(\theta) + 0.05 \times l_{box}(\theta)$ 

 $m_n \leftarrow \beta_1 m_{n-1} + (1 - \beta_1) \nabla loss(\theta)$ 
 $v_n \leftarrow \beta_2 v_{n-1} + (1 - \beta_2) \nabla loss(\theta)^2$ 
 $\hat{m}_n \leftarrow \frac{m_n}{1 - \beta_1}, \hat{v}_n \leftarrow \frac{v_n}{1 - \beta_2}$ 
 $\theta \leftarrow \theta - \frac{\mu}{\sqrt{\hat{v}_n} + \epsilon} \times \hat{m}_n$ 
End
End
Return  $\hat{\theta} = \theta$ 

```

---

$320 \times 320 \times 32$ . The FLOPs of focus and convolution without considering the BN and activate function are shown in (7) and (8), respectively.

$$Focus = 3 \times 3 \times 12 \times 32 \times 320 \times 320 = 353894400. \quad (7)$$

$$Cov = 3 \times 3 \times 3 \times 32 \times 320 \times 320 = 88473600. \quad (8)$$

Based on (7) and (8), the computation of the focus model is about four times larger than convolution. Therefore, using convolution to replace focus can reduce computation.

Based on our proposed ICSP1\_X, ICSP2\_X, SCA, and improved focus modules, the improved YOLOv5s are shown in Figure 2. The pseudo-code of our proposed method is shown in Table 1. The Adam optimizer method is used to optimize the parameters of the network.

## 4. Simulation and Discussion

The experimental environment is configured as follows: the operating system is Ubuntu 18.04. The deep learning framework is Pytorch. Six detection methods that our proposed method, YOLOv3 [10], YOLOv4 [11], YOLOv5s [9], Fast R-Transformer [38], and Mina-Net [31] method are set with the same parameters and pre-training weights are used. The initial learning rate was set to 0.001, and the learning rate was dynamically adjusted using the cosine annealing learning rate. To prevent overfitting, the Adam optimizer is used to adjust the parameters. Pre-training is used in the first five epochs to accelerate the convergence of the model. Pre-training weights are used for all the above methods.

We sourced 195 insulator images from the Power supply and the Internet. The number of defective insulator images is

too small to train the network. Therefore, we use the same method used in the CPLID dataset [39] to expand the data set. We extract defective insulators and fuse them with different backgrounds. The parts of generated defective insulator images are shown in Figure 4. Our proposed defective dataset contains 2300 images. We use 1800 images as training images and 500 images as test images.

**4.1. Ablation Study.** We design a new ResNet1 unit and ResNet2 unit and introduce them into CSP1\_X and CSP2\_X modules to reduce the computation. Besides, we also introduce the channel shuffle operation into improved CSP1\_X and CSP2\_X modules. To simplify, we name the YOLOv5s based on improved CSP1\_X and CSP2\_X without channel shuffle operation as YOLOv5s + Res, YOLOv5s based on improved CSP1\_X and CSP2\_X with channel shuffle operation as YOLOv5s + Res + Cs. Besides, we also design a separate embedded coordinate attention module and introduce it into the improved and CSP2\_X modules. We name the YOLOv5s + Res + Cs based on a separate embedded coordinate attention module as YOLOv5s + Res + Cs + SA. We also name the YOLOv5s + Res + Cs based on the convolutional block attention module [40] as YOLOv5s + Res + Cs + CBAM, and based on Squeeze-and-Excitation [41] as YOLOv5s + Res + Cs + SE. In addition, we also use convolution to replace focus in YOLOv5s. The YOLOv5s + Res + Cs + SA based on a convolution without using focus is named YOLOv5s + Res + Cs + SA-Foc, which is also our complete improved YOLOv5s. The results are shown in Table 2.

The YOLOv5s have the most parameters, FLOPs, and the highest mean average precision (mAP). The YOLOv5s + Res that YOLOv5s based on improved CSP CSP1\_X and CSP2\_X have the smallest parameters, FLOPs, and mAP. As the number of parameters decreases, the mAP also decreases. Although the YOLOv5s + Res + cs that YOLOv5s based on improved CSP CSP1\_X and CSP2\_X with channel shuffle operation have the same number of parameters and FLOPs, but it has larger mAP than YOLOv5s + Res. The channel shuffle operation improves the precision without increasing parameters and FLOPs. Compared with YOLOv5s + Res + Cs, the YOLOv5s + Res + Cs + SE, YOLOv5s + Res + Cs + CBAM and YOLOv5s + Res + Cs + SA that are introduced different attention modules into YOLOv5s + Res + Cs have larger parameters, FLOPs, and mAP. The attention mechanism improves detection precision while increasing parameters and FLOPs. Compared with YOLOv5s + Res + Cs + SE and YOLOv5s + Res + Cs + CBAM, YOLOv5s + Res + Cs + SA has the larger mAP. The SA module is our proposed attention module. This shows that our proposed attention module performs better in precision than SE and CBAM attention modules. Compared with YOLOv5s + Res + Cs + SA, the YOLOv5s + Res + Cs + SA-Foc that uses convolution to replace the focus module to reduce computation has smaller parameters, FLOPs, and mAP.

We propose the improved CSP1\_X module and CSP2\_X module and modify the focus module to reduce computation. We propose an attention module and introduce it into



FIGURE 4: Parts of generated defective insulator images.

TABLE 2: YOLOv5s with different modules.

Methods	Params (M)	FLOPs (G)	mAP (%)
YOLOv5s	7.06	16.3	94.5
YOLOv5s + Res	5.19	12.2	91.7
YOLOv5s + Res + Cs	5.19	12.2	93.2
YOLOv5s + Res + Cs + SE	5.38	12.3	93.5
YOLOv5s + Res + Cs + CBAM	5.40	12.4	93.7
YOLOv5s + Res + Cs + SA	5.47	12.8	94.1
YOLOv5s + Res + Cs + SA-foc(ours)	5.45	12.5	93.6

the network to improve precision. Based on the analysis of Table 2, the improved CSP1\_X module, CSP2\_X module, and modified focus reduce the parameters and FLOPs, and the proposed attention module increases precision.

*4.2. Comparisons for Different Methods.* We compare our complete proposed method with YOLOv3, YOLOv4, and YOLOv5s, Fast R-Transformer, and [38] Mina-Net [31]. We randomly select three different insulators that contain defective insulators. The detection results are shown in Figures 5–7. In each figure, figure (a), figure (b), figure (c), figure (d), figure (e), and figure (f) are obtained by Fast

R-Transformer, Mina-Net YOLOv3, YOLOv4, YOLOv5s, and our proposed method, respectively. In Figure 5, there are two defective insulators, and the YOLOv3 only detects one defective insulator. The YOLOv4, YOLOv5s, Fast R-Transformer, Mina-Net, and our proposed method successfully detect all defective insulators. In Figures 6 and 7, all methods successfully detect all defective insulators. This shows that the proposed method is valid for defective insulator detection.

To verify the performance of the proposed method from a statistical point of view, we use YOLOv3, YOLOv4, YOLOv5s, Fast R-Transformer, Mina-Net, and our proposed to detect all images in the test dataset. The detection results



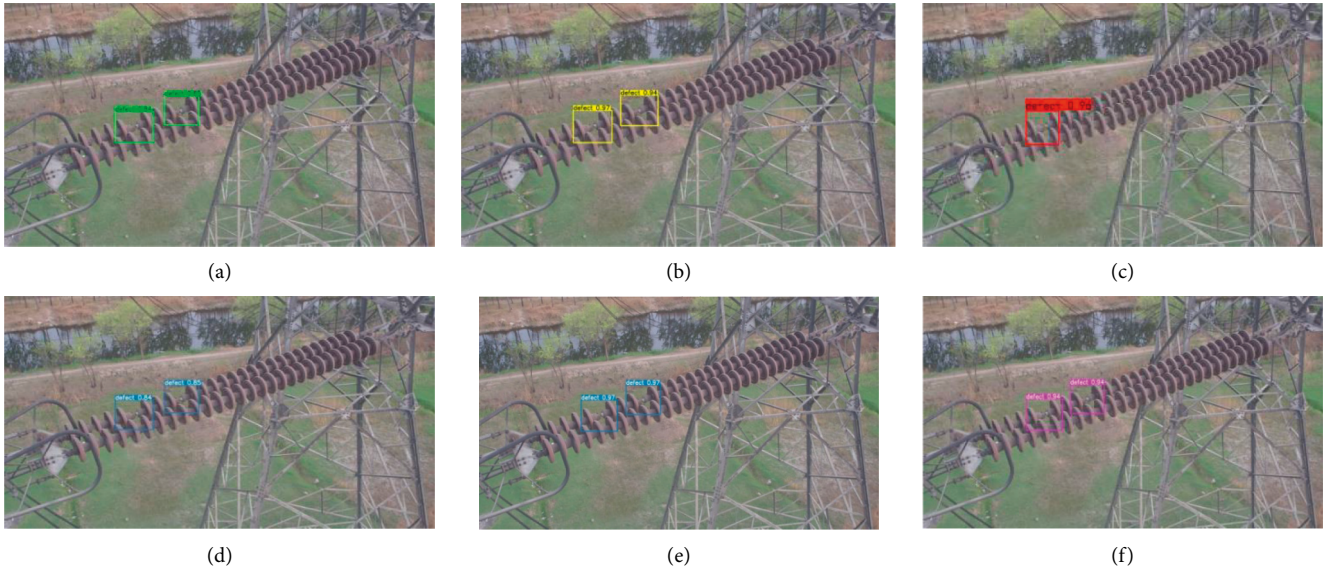


FIGURE 5: Defective porcelain insulator detection.

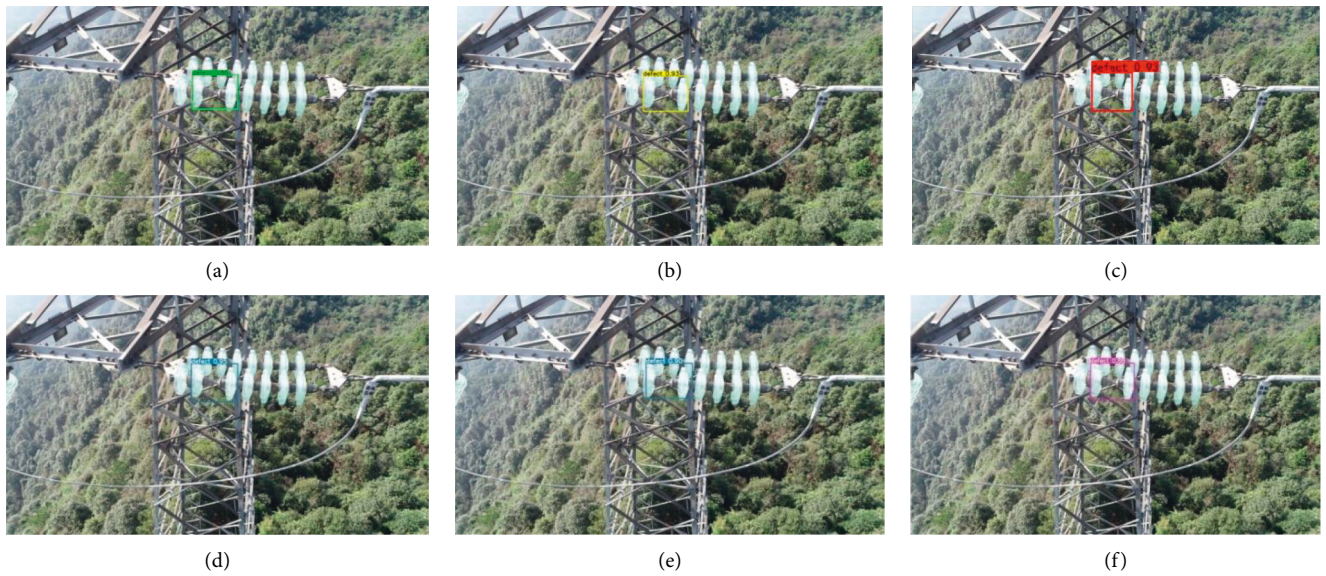


FIGURE 6: Defective toughened glass insulator detection.

are shown in Table 3. The parameters of Fast R-Transformer, Mina-Net, YOLOv3, YOLOv4, YOLOv5s, and our proposed method are 116.10 M, 67.94 M, 61.90 M, 63.94 M, 7.06 M, and 5.45 M, respectively. Our proposed method has the smallest parameters, followed by YOLOv5s. The FLOPs of Fast R-Transformer, Mina-Net, YOLOv3, YOLOv4, YOLOv5s, and our proposed method are 257.5 G, 168.2 G, 155.1 G, 157.1 G, 16.3 G, and 12.5 G, respectively. Our proposed method still has the smallest parameters, followed by YOLOv5s. The mAPs of Fast R-Transformer, Mina-Net, YOLOv3, YOLOv4, YOLOv5s, and our proposed method are 99.7%, 99.2%, 91.5%, 98.7%, 94.5%, and 93.6%, respectively. The Fast R-Transformer has the largest mAP, followed by Mina-Net and YOLOv4. The FPS of Fast

R-Transformer, Mina-Net, YOLOv3, YOLOv4, YOLOv5s, and our proposed method are 38.6, 47.5, 54.6, 52.7, 157.4, and 197.8, respectively. Our proposed method has the largest FPS, followed by YOLOv5s.

Based on the analysis of Table 3, our proposed method has the smallest parameters, FLOPs, and the largest FPS. The FLOPs of different models are computed by model\_into() function in touch\_utils.py. Compared with YOLOv5s, although the mAP of our proposed method is reduced by 0.9%, FPS of our proposed method is increased by about 20%. Although our proposed method has a smaller mAP than other methods, our proposed method has a faster detection speed. Besides, the increase in detection speed is greater than the decrease in precision.

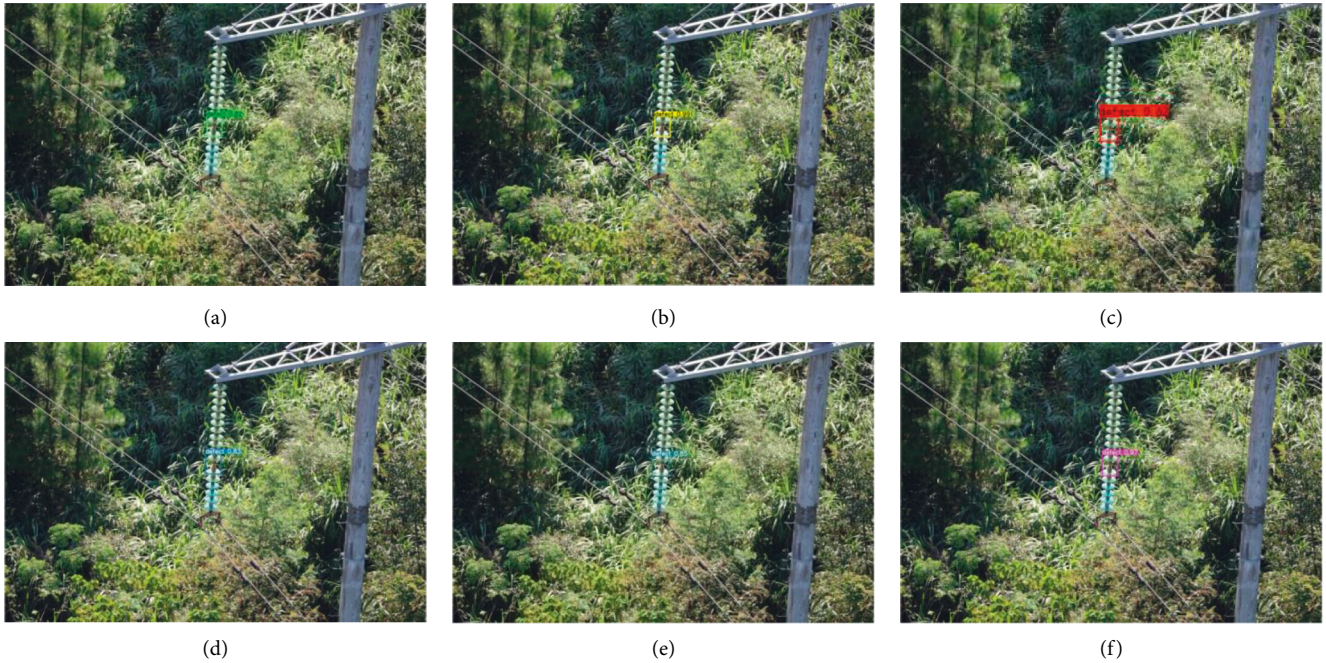


FIGURE 7: Defective suspended glass insulator detection.

TABLE 3: Comparisons for different methods.

Methods	Params (M)	FLOPs (G)	mAP (%)	FPS
Fast R-Transformer	116.1	257.5	99.7	38.6
Mina-net	67.94	168.2	99.2	47.5
YOLOv3	61.90	155.1	91.5	54.6
YOLOv4	63.94	157.1	98.7	52.7
YOLOv5s	7.06	16.3	94.5	157.4
Ours	5.45	12.5	93.6	197.8

## 5. Conclusions

This study proposed a faster defective insulator detection method based on YOLOv5s. It designs two different ResNet units for CSP1\_X module and CSP2\_X module in YOLOv5s to reduce computation, respectively. It also introduced channel shuffle into CSP1\_X module and CSP2\_X module to extract more effective features from different channels without increasing computation. Besides, it designed a separate embedded coordinated attention module and introduced it into the CSP2\_X module to make the network pay more attention to useful information. To reduce the computation, it replaced the focus module using a convolution with stride 2. Compared with defective insulator detection methods based on Fast R-Transformer, Mina-Net, YOLOv3, YOLOv4, and YOLOv5s, our proposed method has the smallest parameters and FLOPs and the largest FPS. This shows that the proposed method has the fastest defective insulator detection speed. Although the other methods except YOLOv3 have larger mAP than our method, the difference in mAP is not large. The mAP of Fast R-Transformer is the largest. Although the mAP of Fast R-Transformer is 6.1% higher than that of our method, the detection speed of Fast R-Transformer is only 1/5 of our

method detection speed. Our method and YOLOv5s have a faster detection speed than others. Compared with YOLOv5s, although the mAP of our proposed method is reduced by 0.9%, the detection speed of our proposed method is increased by about 20%. The increase in detection speed is greater than the decrease in precision. The proposed method has a faster detection speed and an acceptable detection precision.

## Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (61271115) and Science and Technology Research Project of Jilin Provincial Department of Education (JJKH20220054KJ).

## References

- [1] W. Zhao, Q. Dong, and Z. Zuo, "A method combining line detection and semantic segmentation for power line extraction from unmanned aerial vehicle images," *Remote Sensing*, vol. 14, no. 6, Article ID 1367, 2022.
- [2] X. Li, Z. Li, H. Wang, and W. Li, "Unmanned aerial vehicle for transmission line inspection: status, standardization, and perspectives," *Frontiers in Energy Research*, vol. 9, p. 336, 2021.

- [3] R. Xia, Y. Chen, and B. Ren, "Improved Anti-occlusion Object Tracking Algorithm Using Unscented Rauch-Tung-Striebel Smoother and Kernel Correlation Filter," *Journal of King Saud University - Computer and Information Sciences*, 2022.
- [4] J. Zhang, W. Feng, T. Yuan, J. Wang, and A. K. Sangaiah, "SCSTCF: spatial-channel selection and temporal regularized correlation filters for visual tracking," *Applied Soft Computing*, vol. 118, Article ID 108485, 2022.
- [5] Y. Chen, L. Liu, V. Phonevilay et al., "Image super-resolution reconstruction based on feature map attention mechanism," *Applied Intelligence*, vol. 51, no. 7, pp. 4367–4380, 2021.
- [6] L. Zhao, Y. Zhang, and Y. Cui, "An attention encoder-decoder network based on generative adversarial network for remote sensing image dehazing," *IEEE Sensors Journal*, vol. 22, no. 11, pp. 10890–10900, 2022.
- [7] W. Rahmaniari and A. Hernawan, "Real-time human detection using deep learning on embedded platforms: a review," *Journal of Robotics and Control (JRC)*, vol. 2, no. 6, pp. 462–468, 2021.
- [8] H. Mei, H. Jiang, F. Yin, L. Wang, and M. Farzaneh, "Terahertz imaging method for composite insulator defects based on edge detection algorithm," *IEEE Transactions On Instrumentation And Measurement*, vol. 70, pp. 1–10, 2021.
- [9] J. Glenn, "YOLOv5," 2020, <https://github.com/ultralytics/YOLOv5>.
- [10] J. Redmon and A. Farhadi, "Yolov3: an incremental improvement," 2018, <http://arxiv.org/abs/1804.02767>.
- [11] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: optimal speed and accuracy of object detection," 2020, <http://arxiv.org/abs/2004.10934>.
- [12] Z. Zhang, Z. Xiong, B. Zhang, Y. Yang, and E. Fu, "Detection for small target ship in remote sensing image based on super resolution reconstruction technology," *Journal of Northeast Electric Power University*, vol. 42, no. 2, pp. 33–40, 2022.
- [13] K. He, G. Gkioxari, P. Dollár, R. Girshick, and R.-C. N. N. Mask, *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969, 2017.
- [14] Z. Cai and N. Vasconcelos, "Cascade R-CNN: delving into high quality object detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6154–6162, 2018.
- [15] P. Sun, R. Zhang, Y. Jiang et al., "End-to-End object detection with learnable proposals," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14454–14463, Nashville, TN, USA, 2021.
- [16] W. Liu, D. Anguelov, D. Erhan et al., *SSD: Single Shot Multibox Detector, European Conference on Computer Vision*, pp. 21–37, Springer, Berlin/Heidelberg, Germany, 2016, <http://arxiv.org/abs/1512.02325>.
- [17] J. Jeong, H. Park, and N. Kwak, "Enhancement of SSD by concatenating feature maps for object detection," 2017, <http://arxiv.org/abs/1705.09587> arXiv preprint arXiv:1705.09587.
- [18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788, Las Vegas, NV, USA, June 2016.
- [19] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7263–7271, Honolulu, HI, USA, July 2017.
- [20] Y. Yang and H. Deng, "GC-YOLOv3: you only look once with global context block," *Electronics*, vol. 9, no. 8, p. 1235, 2020.
- [21] Z. Qu, F. Zhu, and C. Qi, "Remote sensing image target detection: improvement of the YOLOv3 model with auxiliary networks," *Remote Sensing*, vol. 13, no. 19, p. 3908, 2021.
- [22] Y. Yin, H. Li, and W. Fu, "Faster-YOLO: an accurate and faster object detection method," *Digital Signal Processing*, vol. 102, Article ID 102756, 2020.
- [23] K. Cai, X. Miao, W. Wang, H. Pang, Y. Liu, and J. Song, "A modified YOLOv3 model for fish detection based on MobileNetv1 as backbone," *Aquacultural Engineering*, vol. 91, Article ID 102117, 2020.
- [24] T. Deng and Y. Wu, "Simultaneous vehicle and lane detection via MobileNetV3 in car following scene," *Plos one*, vol. 17, no. 3, Article ID e0264551, 2022.
- [25] C. Y. Wang, A. Bochkovskiy, and H. Y. M. Liao, "Scaled-yolov4: scaling cross stage partial network," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13029–13038, 2021.
- [26] F. Guo, Y. Qian, and Y. Shi, "Real-time railroad track components inspection based on the improved YOLOv4 framework," *Automation in Construction*, vol. 125, Article ID 103596, 2021.
- [27] J. Yu and W. Zhang, "Face Mask wearing detection algorithm based on improved YOLO-v4," *Sensors*, vol. 21, no. 9, p. 3263, 2021.
- [28] Z. Ma, Y. Li, M. Huang, Q. Huang, J. Cheng, and S. Tang, "A lightweight detector based on attention mechanism for aluminum strip surface defect detection," *Computers in Industry*, vol. 136, 103585.
- [29] C. Liu, Y. Wu, J. Liu, Z. Sun, and H. Xu, "Insulator faults detection in aerial images from high-voltage transmission lines based on deep learning model," *Applied Sciences*, vol. 11, no. 10, p. 4647, 2021.
- [30] Z. Qiu, X. Zhu, C. Liao, D. Shi, and W. Qu, "Detection of transmission line insulator defects based on an improved lightweight YOLOv4 model," *Applied Sciences*, vol. 12, no. 3, p. 1207, 2022.
- [31] H. He, X. Huang, Y. Song et al., "An insulator self-blast detection method based on YOLOv4 with aerial images," *Energy Reports*, vol. 8, pp. 448–454, 2022.
- [32] Z. Feng, L. Guo, D. Huang, and R. Li, "Electrical Insulator Defects Detection Method Based on YOLOv5," in *Proceedings of the 2021 IEEE 10th Data Driven Control and Learning Systems Conference (DDCLS)*, pp. 979–984, Suzhou, China, May 2021.
- [33] Y. Lan and W. Xu, "Insulator defect detection algorithm based on a lightweight network," *Journal of Physics: Conference Series*, vol. 2181, no. 1, Article ID 012007, 2022.
- [34] Y. Wang, P. Cao, X. Wang, and X. Yan, "Research on insulator self explosion detection method based on deep learning," *Journal of Northeast Electric Power University*, vol. 40, no. 3, pp. 33–40, 2020.
- [35] M. F. Palangar, S. Mohseni, M. Mirzaie, and A. Mahmoudi, "Designing an automatic detector device to diagnose insulator state on overhead distribution lines," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 2, pp. 1072–1082, 2022.
- [36] B. Wang, M. Dong, M. Ren et al., "Automatic fault diagnosis of infrared insulator images based on image instance segmentation and temperature analysis," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 8, pp. 5345–5355, 2020.
- [37] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13713–13722, 2021.

- [38] S. Dian, X. Zhong, and Y. Zhong, "Faster R-Transformer: an efficient method for insulator detection in complex aerial environments," *Measurement*, vol. 199, Article ID 111238, 2022.
- [39] X. Tao, D. Zhang, Z. Wang, X. Liu, H. Zhang, and D. Xu, "Detection of power line insulator defects using aerial images analyzed with convolutional neural networks," *IEEE Transactions On Systems, Man, and Cybernetics: Systems*, vol. 50, no. 4, pp. 1486–1498, 2020.
- [40] S. Woo, J. Park, J. K. Leeand, and S. Kweon, "CBAM: convolutional block Attention module," *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3–19, 2018.
- [41] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141, 2018.