

Received September 27, 2020, accepted October 5, 2020, date of publication October 9, 2020, date of current version October 21, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3029857

# Detection of Insulator Defects With Improved ResNeSt and Region Proposal Network

SHUQING WANG<sup>1</sup>, YIFAN LIU<sup>1</sup>, YIHUI QING<sup>1</sup>, CHENXI WANG<sup>1</sup>,  
TIANZE LAN<sup>1</sup>, AND RUOTIAN YAO<sup>2</sup>

<sup>1</sup>Hubei Power Grid Intelligent Control and Equipment Engineering Technology Research Center, Hubei University of Technology, Wuhan 430068, China

<sup>2</sup>School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, China

Corresponding author: Yifan Liu (embalming@foxmail.com)

This work was supported by the National Natural Science Foundation of China (Research on Stochastic Game and Optimization of Complex Smart Grid Based on Multi Agent) under Grant 51407063.

**ABSTRACT** Insulator is an important part of transmission line. Defective insulators will cause potential safety hazard to transmission lines. Image detection technology can improve the efficiency of insulator defect detection and greatly reduce the maintenance cost. However, the existing insulator defect detection technology has the disadvantages of low accuracy and long detection time. An insulator defect detection method based on improved ResNeSt and Region Proposal Network (RPN) was proposed. First, this method builds a new network based on ResNeSt. Secondly, we added the improved RPN to the improved ResNeSt for feature extraction, to better detect minor defects on insulators. Finally, we enhanced the data processing and labeled the open insulator data set. On this data set, the proposed model is tested and a large number of controlled experiments are done. The results show that the proposed network is more accurate and faster than the control group. Moreover, the proposed network has an accuracy rate of 98.38% for insulator defect detection, which can detect 12.8 pictures per second. The proposed method has good efficiency and practicability in aerial photo insulator defect detection.

**INDEX TERMS** Computer vision, defect detection, deep learning, residual network, region proposal.

## I. INTRODUCTION

Insulator is an important part of transmission line, whose main function is electrical insulation and line support. When problems such as break, crack and dirt occur to the insulator, the insulator is prone to breakdown, which results in zero insulation resistance at both ends of the insulator string. The insulation of the insulator is lost, resulting in the interruption of power supply, which will lead to a blackout. To check and maintain insulators regularly to ensure the safety of reliable power supply system, insulator defect detection has become an important issue [1].

In recent years, machine vision methods such as Histogram of Oriented Gradient (HOG) [2] and Local Binary Patterns (LBP) [3] are often used in insulator defects detection. Compared with the original manual detection method, these methods have the advantages of fast detection speed and low cost. However, due to the fact that the actual working insulator is often in a complex background, it is extremely easy to be affected by light and noise [4]. The application of

these methods is not good in practice, and it is easy to identify the shadow as a defect.

Compared with the original methods of insulator defects detection which are based on machine vision, the methods based on deep learning can extract the image features efficiently and automatically, which greatly improves the efficiency and accuracy of defects detection. The methods based on deep learning have attracted the attention of many experts in related fields.

In general, most of the methods which are based on artificial features and machine learning are sensitive to complex background interference. Most of these methods are time-consuming and far from real-time applications. Most importantly, neither the existing feature-based methods, machine learning methods or deep learning methods have systematically analyzed and solved the problem of insulator multi-defect detection [5]. Therefore, it is meaningful to propose a method that can solve the existing problems.

In order to ensure the accuracy and speed of insulator defects detection, an improved ResNeSt [6] and Region Proposal Network (RPN) [7] is proposed. This proposed method can identify three kinds of defects in four kinds of

The associate editor coordinating the review of this manuscript and approving it for publication was Szidónia Lefkovits<sup>1</sup>.

common insulators. Through the experiments on the data set, this method is more accurate and faster than other insulator defects detection methods based on deep learning. This paper mainly overcomes three challenges:

- a) Because ResNeSt lacks practical application, it is necessary to improve the data set before training the model.
- b) At present, deep learning is mainly used in the direction of object recognition, however, the effect is slightly poor when detecting irregular and irregular objects such as defects. Therefore, it is necessary to optimize the network to obtain better performance.
- c) In aerial photography, the insulator in the working state often only occupies a small part in the high-resolution image. The convolution detection method is very slow, which cannot meet the requirements of real-time detection. And small object detection has always been an important difficulty in object recognition direction. Therefore, it is necessary to optimize the speed and accuracy of model detection.

The rest of this paper is organized as follows. Section 2 reviews the previous research on insulator defects detection, and introduces the relevant methods used in this paper. Section 3 introduces the characteristics of insulator defects, describes the overall structure of the network, and improves the original ResNeSt and RPN. Section 4 evaluates the proposed network and compares it with other networks. Finally, section 5 gives the conclusion of our proposed network and the direction of our future work.

## II. RELATED WORKS

In this section, we review the previous research on insulator defects detection.

### A. INSULATOR DEFECTS DETECTION

Traditional machine learning methods mainly locate obvious defects through HOG and LBP features. Wu *et al.* [8] constructed an insulator data set and calculated the HOG features of each image. On this foundation, Principal Component Analysis (PCA) was used in this paper, a complete set of data set is established and a classifier based on sparse representation is trained to obtain the location of insulation defects. However, in the complex background of texture interference, only using the over complete HOG features set to train the classifier cannot achieve good performance. In order to offset his shortcomings, Tiantian *et al.* [9] used HOG and LBP features, which can form a fusion feature. This special fusion feature can train Support Vector Machine (SVM) classifier. Through rigorous experiments, they found that this method can achieve multi-angle insulator location in complex scene. Zuo *et al.* [10] combines Haar-like features, integral image features and HOG features to train cascade classifiers and SVM classifiers. Then the two classification models are applied to locate insulators. Finally, the defects location is determined by incremental contour value method. Oberweiger *et al.* [11] proposed a K-Nearest Neighbor (KNN) classifier to distinguish insulator from background clutter,

and developed an automatic insulator defects detector based on elliptic Fourier descriptor to analyze the defects of each insulator. It is possible to detect multiple insulator defects by their proposed method. However, there is no public data set for insulator defects detection. The data set they built only contains 10 images of insulator defects. Although the methods based on machine learning improve the accuracy of insulator location and defects detection, these methods have the common limitation of time-consuming because the sliding window strategy must be used to detect the whole aerial image.

In recent years, some scholars take the deep learning network as the feature extraction method of insulator defects detection. Compared with the original machine vision feature extraction method, deep learning can accurately locate the defect location of insulator under the natural complex background. Even in the bad natural environment such as shadow, it can still better locate the boundary of defect. Hu [12], Zhao [13] used Fast RCNN and Faster RCNN to locate the insulator, but the training process of Fast RCNN and Faster RCNN is complex and difficult to deploy, so it is unable to locate insulator in real time in aerial images. Ling [14] used Faster RCNN which uses rectangular bounding box to mark the insulator position in aerial image. U-net is developed to segment the defect contour in the rectangular boundary box. When the insulators overlap seriously, the performance of this string structure will be reduced. In addition, there is no common data set for insulator defects detection, so it is difficult to train a good performance end-to-end network to detect insulator defects. Wu *et al.* [15] designed an insulator defect location method based on Region of Interest (ROI) by improving YOLOv3. Although this method does improve the quality of insulator defect detection in most aerial images, it is still not a real-time solution. In addition, there is no public data set for insulator defect detection. Therefore, it is difficult to train a good performance insulator defect detection network. To meet this challenge, they segmented insulator strings from aerial images. They pasted segmented insulator strings onto aerial images containing only background to enhance their insulator defect data set. However, the insulator defects in the simulated aerial images are similar to those in the original aerial images. This shortcoming affects the experimental results and the generalization ability of network performance. Prates *et al.* [16] shot a single insulator in a simple background as the network pre-training data set, and set up the insulator in the outdoor simulation work as the real insulator data set. They trained with the original deep learning network, without further optimization of the network performance.

There are few researches on insulator defect detection through deep learning, which is because the number of defective insulators in normal working environment is very small, which leads to the collection of defective insulator data sets taking a long time. In this paper, based on the unlabeled real defective insulator data sets published by Prates, we carry out defect labeling, data set enhancement and adaptation to

ResNeSt network. The following section describes in detail how we improve.

### B. RESIDUAL BLOCK

The network model with residual block, such as ResNet [17], is obviously better than the original convolution network in the field of image classification. With the increase of network layers, the network degrades. And with the increase of network layers, the loss of training set decreases gradually, and then tends to saturation. When the depth of network is increased, the loss of training set will increase. After the introduction of residual block, the network can reach a very deep level, and the effect of the network will become better because ResNet has a deeper convolution layer. Due to the down sampling effect of original convolution neural network, small objects cannot obtain significant features. At the same time, residual learning method is used to combine deep feature mapping with shallow feature mapping. The algorithm effectively takes advantage of the combination of high-level features and low-level features, and can better adapt to the detection of insulator like small targets in high-resolution images.

Xie *et al.* [18] proposed ResNeXt which is the upgraded version of ResNet. He thinks that the original deep learning network to improve the accuracy of the model is to deepen or widen the network, but with the increase of the number of hyperparameters, such as the number of channels, filter size, etc., the difficulty of network design and calculation cost will also increase. Therefore, the structure of ResNeXt can improve the accuracy without increasing the complexity of parameters and reduce the number of hyperparameters. The core innovation of ResNeXt is to propose aggregated transformations, which uses parallel stacking the same topology blocks to replace the three-layer convolution block of the original ResNet, which improves the accuracy of the model without significantly increasing the magnitude of parameters. At the same time, due to the same topology structure, the number of hyperparameters is also reduced, which is convenient for model transplantation.

Zhang *et al.* proposed ResNeSt which was an upgraded version of ResNeXt. ResNeSt combines the group convolution of ResNeXt and the channel attention mechanism of SE-Net. ResNeSt groups the channels, using the channel attention mechanism for each group, and retaining the residual structure of ResNet. ResNeSt surpasses its predecessors ResNet, ResNeXt, SE-Net and EfficientNet in image classification. The mAP of Faster RCNN with ResNeSt50 as the backbone was 3.08% higher than that with ResNet50. Using ResNeSt50 as the backbone, the mIoU (mean Intersection over Union) of DeeplabV3 is 3.02% higher than using ResNet50.

### C. REGION PROPOSAL NETWORK

Region Proposal Network (RPN) is proposed in Faster RCNN. Because of introducing multi-scale sliding window to traverse each spatial position of feature graph, RPN greatly

improves the recall rate of object detection. However, RPN only extracts candidate objects from a certain depth convolution feature layer, and its fixed size convolution kernel limits the size of visual receptive field of a single feature layer. Therefore, Feature Pyramid Networks (FPN) generates multi-scale candidate targets on multiple feature layers, which further improves the recall rate of object detection [19]. Based on that, this paper carried out the experimental analysis of the defect recall rate of multi-scale RPN insulator, and found that the performance of different depth convolution feature layer on different scale insulator defect candidate target recall rate has great difference. Large size insulator defects have higher recall rate in high-level feature, while small size insulator defects have higher recall rate in low-level features with high-resolution. Therefore, according to the effective receptive field size of each depth convolution feature layer, this paper adopts the scale complementary strategy to divide the candidate target into three paths of RPN to adapt to the multi-scale variation of insulator defects.

## III. INSULATOR DEFECTS AND DETECTION NETWORK

In this section, we analyze the characteristics of insulator defects and propose a defect detection network based on ResNeSt. We will introduce the overall architecture and detailed core components of the improved ResNeSt.

### A. INSULATOR DEFECTS

The common insulators are Polymeric Grey Insulator (PGI), Ceramic Pin Insulator (CPI), Glass Green Insulator (GGI) and Ceramic Bicolor Insulator (CBI). The common insulator defects are break, crack and dirty. First of all, break means that most of the insulators are damaged and lose their working capacity. Crack usually only accounts for a small part of the insulator, but the damage often makes the insulator lose most of the insulator. In rain and fog weather, the insulation performance drops sharply, resulting in flashover or insulation breakdown, which will result in grounding fault. Dirt refers to the pollution material with conductive performance accumulated on the surface of line insulator, which will greatly reduce the insulation level of insulator after being affected by moisture in wet weather, and the flashover accident occurs under normal operation. Secondly, different types of insulators show different characteristics on different defects, including color features, shape features and regional characteristics. We need to design a reasonable network for these features.

### B. IMPROVED RESNEST

ResNeSt is a kind of network based on ResNet and introduced split attention block proposed by Zhang *et al.* The network achieves 81.13% top-1 accuracy rate on ImageNet. The performance is significantly improved but the number of parameters is not significantly increased. It has great application value. Therefore, this network is considered to complete the insulator defect detection. ResNeSt introduces Split-Attention block on ResNet. The Split-Attention

block allows features map between different feature mapping groups. It is a computing unit composed of feature mapping module and split attention operation module. ResNeSt can divide the input insulator characteristic diagram into the number of groups determined by the hyperparameter  $K$ , and introduces a new cardinal parameter  $R$ , and  $R = G/K$ , where  $G$  is the total number of essential factor groups. Each individual group is then mapped to  $\{L_1, L_2, \dots, L_G\}$ . Next, the elements of each cardinal group are summed and fused by multiple partitions. The combination of each cardinal group can be obtained, which is represented by  $\hat{L}^k$ , where  $\hat{L}^k \in \mathbb{R}^{W \times H \times C/K}$ , and  $\{k \in 1, 2, \dots, K\}$ ,  $W, H$  and  $C$  are the output feature mapping sizes of the segmented attention module.

Considering that the defective insulator in aerial photos may account for a small part of the whole picture. And the resolution is often very low. In order to speed up the calculation of ResNeSt, as shown in the middle part of Fig. 1. The following improvements are made.

We proposed a thinning algorithm based on ResNeSt to refine the detection effect of insulator defects adaptively. As shown in Fig. 1, the feature image split by ResNeSt block is introduced as a fusion network to obtain the final segmentation result. In order to refine the characteristics of insulators and defects, the improved ResNeSt module generates table transformed local propagation coefficient mapping for all positions. The formula is as follows:

$$W_i^p = \frac{\exp(h_i^p)}{\sum_{t=1}^{m \times m} \exp(h_t^p)}, \quad p \in 1, 2, \dots, m \times m \quad (1)$$

where  $h_i^p$  is the confidence level of neighborhood  $p$  at location  $i$ , and  $m \times m$  is the size of propagation neighborhood. Finally, the final output is obtained through the following processing:

$$g_i = \sum_{p=1}^{m \times m} w_i^p \cdot f_i^p \quad (2)$$

Among them,  $f_i^p$  is the confidence vector of neighborhood  $p$  at position  $i$  of the improved ResNeSt module,  $g_i$  is the final prediction vector of position  $i$ .

Like the standard residual block, if the input and output feature maps share the same shape, the final output of the improved ResNeSt can be expressed as  $y_1 = h(x_l) + \mathbb{F}(x_l, W_l)$ , where  $f(y_l) = x_{l+1}$  and  $f(y_l)$  is the activation functions of  $y_l$ ,  $h(x_l)$  is the direct mapping of  $x_l$ , and  $\mathbb{F}(x_l, W_l)$  is the residual part. For a deeper layer  $L$ , the relationship between it and layer  $l$  can be expressed as follows:

$$x_L = x_l + \sum_{i=l}^{L-1} \mathbb{F}(x_i, W_i) \quad (3)$$

This formula proves two properties of the improved ResNeSt:

- a) The layer  $L$  can be expressed as the sum of any layer  $l$  shallower than it and the residual part between them.
- b) When  $l = 0$ ,  $x_L = x_0 + \sum_{i=0}^{L-1} \mathbb{F}(x_i, W_i)$  means that  $l$  is the sum of the units of each residual block feature.

According to the derivative chain rule in Backpropagation algorithm, the gradient of loss function  $\varepsilon$  with respect to  $x_l$  can be expressed as follows:

$$\begin{aligned} \frac{\partial \varepsilon}{\partial x_l} &= \frac{\partial \varepsilon}{\partial x_L} \frac{\partial x_L}{\partial x_l} = \frac{\partial \varepsilon}{\partial x_L} \left( 1 + \frac{\partial}{\partial x_l} \sum_{i=l}^{L-1} \mathbb{F}(x_i, W_i) \right) \\ &= \frac{\partial \varepsilon}{\partial x_L} + \frac{\partial \varepsilon}{\partial x_L} \frac{\partial}{\partial x_l} \sum_{i=l}^{L-1} \mathbb{F}(x_i, W_i) \end{aligned} \quad (4)$$

This formula proves two other properties of the improved ResNeSt:

- a) In the whole training process,  $\partial(\sum_{i=l}^{L-1} \mathbb{F}(x_i, W_i))/\partial x_l$  cannot always be -1, that is to say, there will be no gradient disappear in the residual network.
- b)  $\partial \varepsilon / \partial x_L$  means that the gradient of the layer  $L$  can be transferred directly to any layer shallower than it.

By analyzing the forward and backward processes of the improved ResNeSt, it is found that when the residual block satisfies the above two assumptions, the information can be transmitted smoothly between the high level and the low level, which indicates that these two assumptions are sufficient conditions for the improved ResNeSt to train the depth model.

C. MULTI-SCALE RPN

Small defects are common in insulator defects. Because some insulator defects are far away from the camera, the target size is small, which makes the insulator defect target occupy a small pixel in the image, and the corresponding area contains less information. Therefore, it is easy to miss

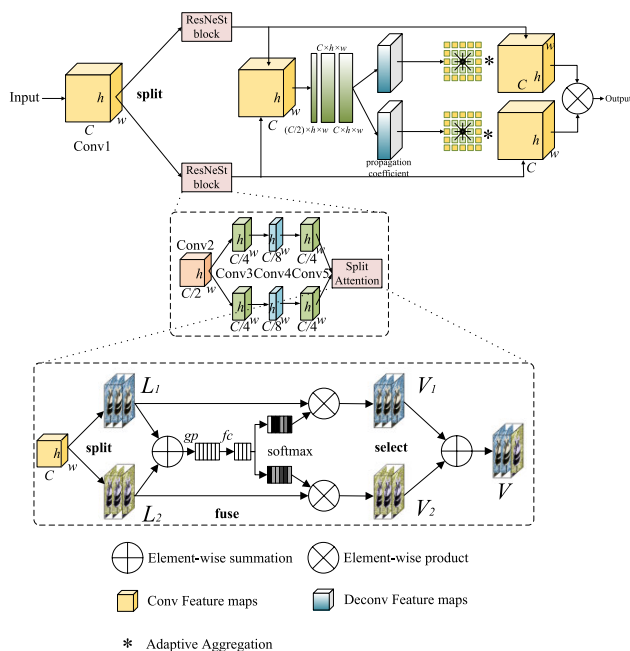


FIGURE 1. Improved ResNeSt for insulator defect detection.



the detection, which affects the detection accuracy of the algorithm. Therefore, it is difficult to identify and locate the small-scale insulator defect target in the field of target detection. In order to solve the difficult problem of small object detection in challenge c, ResNeSt is combined with multi-scale RPN. We use ResNeSt whose backbones are ResNet50 and ResNet101 as the basic network of feature extraction. Conv3, Conv4 and Conv5 are defined to represent the last residual block ResNeSt\_3d, ResNeSt\_4f and ResNeSt\_5c in each stage of the infrastructure network. For the three branches, the height (pixel value of insulators instance height) of the effective real annotation box in RPN is within the range of  $[\text{inf}, 50]$ ,  $[\text{inf}, \text{inf}]$  and  $[100, \text{inf}]$ . The ground truth annotation across this range is regarded as invalid annotation and does not participate in the training of the RPN branch.

The RPN is trained with different target paths due to different RPN scales. The RPN multi task loss function is defined as follows.

$$Loss_{RPN} = Loss_1 + \varphi[\tau = 1]Loss_2 \quad (5)$$

where  $Loss_1$  is the classification loss using cross entropy loss function.  $Loss_2$  is the position regression loss using Smooth L1 loss function [20].  $\varphi$  is a hyperparameter.  $\tau = 1$  means that only positive samples are used for position regression. In order to shorten the high-resolution features of the graph to the low-level features, the high-resolution features of the graph are aggregated. In the bottom-up feature coding path, the max-pooling method is used to realize down sampling. The purpose is to reduce the parameters and keep the invariance of rotation and translation. Max-pooling is to take the maximum value of adjacent features and retain more texture feature information. The average pooling can reduce the variance error of the estimated value caused by the size limitation of the neighborhood, and more emphasis on the lower sampling of the overall feature information. At the same time, it is more conducive to information transfer to the next feature layer. According to the effectiveness of different resolution feature layers for different scale insulator defects, the candidate region feature codes are extracted by multi-scale detection method combined with multi-scale insulator defect candidate set  $C_i = \{C_s, C_a, C_l\}$  generated by multi-path RPN and aggregate feature  $Q_i = \{Q_3, Q_4, Q_5\}$  obtained by cross-scale aggregation feature network module. Firstly, the candidate regions of insulator defects in  $C_i$  set are generated by the main detection branch of multi-path RPN to match the corresponding aggregation feature  $Q_i$  generated by cross-scale aggregation feature network. The region of interest of the feature layer is obtained. Then,  $(C/8) \times w \times h$  feature is obtained by using the ROI-pooling normalized extracted feature coding. Then, the extracted feature codes are transformed from the full connection layer to the high-dimensional feature vectors, and the confidence scores and four coordinate offsets of the candidate regions are accurately calculated to obtain the final detection results. The other two auxiliary detection branches are similar. For the candidate regions of

different scale sets, the corresponding detection branches are used. Each detection branch training has a real class label  $x_{ml}^*$  and a real label box  $box^*$ . The loss function of single branch insulator defect detection training is defined as follows.

$$Loss(x_{ml}, box) = Loss_1(x_{ml}, x_{ml}^*) + \omega x_{ml}^* Loss_3 \times (box, box^*) \quad (6)$$

where  $Loss_3$  is the regression loss function of candidate target and  $Loss_3(box, box^*) = \mathbb{R}(box - box^*)$ .  $\mathbb{R}$  is Smooth L1 loss function.  $x_{ml}$  is the confidence score of network candidate target frame.  $box$  is the predicted candidate target frame.  $\omega$  is the loss function of balancing classification and regression tasks. When the overlap degree between the predicted candidate target frame and any real annotation frame is greater than the constant  $\lambda$  ( $0 < \lambda < 1$ ), then  $p^* = 1$ , otherwise  $p^* = 0$ . The proposed multi-scale RPN based on improved ResNeSt is shown in Fig. 2. The implementation process of the improved multi-scale RPN is shown in Algorithm 1.

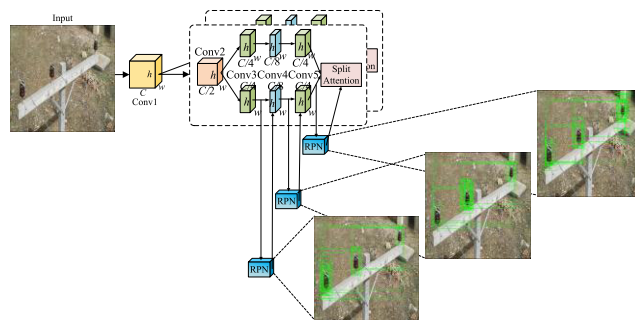


FIGURE 2. Improved RPN for insulator defect detection. Improved RPN activates two parts split by the Split-attention block at the same time.

#### IV. EXPERIMENTS AND DISCUSSION

The experiment is based on PyTorch 1.5.1 in Python 3.7. PyTorch is a library specially built for deep learning model. The above experiments were performed on Intel (R) Xeon (R) gold CPU 6148@3.7GHz 32GB, GPU run on NVIDIA Tesla V100 32GB and Ubuntu 16.04 LTS.

##### A. DATA ANNOTATION

This data set was a defective insulator data set published by Prates *et al.* in 2019. The data set contains four common 15kV distribution insulators, which are Polymeric Grey Insulator (PGI), Ceramic Pin Insulator (CPI), Glass Green Insulator (GGI) and Ceramic Bicolor Insulator (CBI). They were respectively installed on a teaching high-voltage cable. A total of 2560 real photos of defective insulator and normal insulator were taken. There are too few images in the public data set. In order to balance the number of defect images, optimize the data set and enhance the data set, we remove some unclear and ambiguous images. Through data enhancement techniques such as rotation, brightness adjustment and translation, the number of data sets was increased to 48000 photos.

**Algorithm 1** Improved Multi-Scale RPN

**Input:** training sample  $U = \{(xml_i^*, box_i^*)\}_{i=1}^N$ , iterations  $K$ , learning rate  $\xi_\epsilon$  and  $\xi_{\epsilon_b}$  and minibatch  $B$

**Output:** weight  $\epsilon$  and  $\epsilon_b$

- 1 **for**  $t = 1$  to  $T$  **do**
- 2   **for** minibatch **do**
- 3     **if**  $box_h^* < 50$
- 4        $Loss_{far} = Loss_1 + \varphi[\tau = 1]Loss_2$
- 5        $Loss_{far}(xml, box) = Loss_1(xml, xml^*) + \omega xml^* Loss_3(box, box^*)$
- 6     **if**  $50 \leq box_h^* < 100$
- 7        $Loss_{medium} = Loss_1 + \varphi[\tau = 1]Loss_2$
- 8        $Loss_{medium}(xml, box) = Loss_1(xml, xml^*) + \omega xml^* Loss_3(box, box^*)$
- 9     **if**  $box_h^* \geq 100$
- 10        $Loss_{near} = Loss_1 + \varphi[\tau = 1]Loss_2$
- 11        $Loss_{near}(xml, box) = Loss_1(xml, xml^*) + \omega xml^* Loss_3(box, box^*)$
- 12      $\epsilon = \text{SGD}(\nabla_{\epsilon}(Loss_1 + \omega xml^* Loss_3), \epsilon, \xi_\epsilon)$
- 13      $\epsilon_b = \text{SGD}(\nabla_{\epsilon_b}(Loss_1 + \omega xml^* Loss_3), \epsilon_b, \xi_{\epsilon_b})$
- 14   **end**
- 15 **end**
- 16 **return**  $\epsilon, \epsilon_b$

There are 12000 photos for each type of insulator. Each type of insulator contains 4320 defective insulator images and 7680 normal insulator images. The resolution of the above data sets is  $224 \times 224$ . After sorting out, it is found that there are three kinds of insulator defects in each insulator data set, which are break, crack and dirt. For each type of insulator, there are 1440 pictures of each of the three defects. Public data sets are unlabeled. We use LabelImg [21] image annotation tool to label the insulator and the defects on the insulator. The tool will generate an xml file, and each xml file corresponds to each image one by one. Each xml file contains the boundary box coordinates of insulator location, insulator type, boundary box coordinate of defect location and defect type.

**B. EVALUATION METHOD**

In order to comprehensively and objectively evaluate the performance of the proposed method, we use the following indicators for comprehensive evaluation [22].

**ACC (Accuracy):**  $ACC = (TP + TN) / (TP + FP + TN + FN)$ . TP (True Positions) is the number of positive samples correctly identified as positive samples. FP (False Positions) is the number of negative samples wrongly identified as positive samples. TN (True Negatives) is the number of negative samples correctly identified as negative samples. FN (False Negatives) is the number of positive samples that are wrongly identified as negative samples.

**Mean Average Precision (mAP):** Average Precision (AP) is the area under the precision recall curve. Where precision =  $TP / (TP + FP)$ , recall =  $TP / (TP + FN)$ .

**TABLE 1.** Detection results on insulator data set.

Methods	Backbone	ACC	mAP (%)	AUC	FPS
YOLOv3	Darknet-53	0.7887	70.6	0.7228	19.23
YOLOv3 by Tao	Darknet-53	0.8350	74.3	0.7566	23.26
YOLOv4	CSPDarknet-53	0.8562	75.6	0.7854	<b>25.21</b>
Faster RCNN by Ling	VGG16	0.8825	77.8	0.8311	9.09
Faster RCNN	ResNet50	0.8388	74.4	0.8153	8.33
Faster RCNN	ResNet101	0.8738	76.9	0.8291	11.11
Faster RCNN	ResNet50-RPN	0.8863	80.0	0.8793	3.23
Faster RCNN	ResNet101-RPN	0.9262	83.3	0.8856	3.70
Faster RCNN	ResNeSt50	0.9041	82.3	0.8856	12.30
Faster RCNN	ResNeSt101	0.9342	85.7	0.9128	7.55
Faster RCNN	ResNeSt50-RPN	0.9453	91.0	0.9324	9.36
Faster RCNN	ResNeSt101-RPN	0.9771	93.3	0.9541	5.23
Faster RCNN	proposed method	0.9838	95.8	0.9763	12.80
RetinaNet	ResNet50	0.8858	79.9	0.8956	12.50
RetinaNet	ResNet101	0.8993	80.5	0.9032	9.09
RetinaNet	ResNet50-RPN	0.9301	84.9	0.9225	10.00
RetinaNet	ResNet101-RPN	0.9398	85.2	0.9385	2.38
RetinaNet	ResNeSt50	0.9543	87.5	0.9243	11.43
RetinaNet	ResNeSt101	0.9643	90.1	0.9373	9.17
RetinaNet	ResNeSt50-RPN	0.9732	94.6	0.9543	6.46
RetinaNet	ResNeSt101-RPN	0.9814	95.9	0.9713	5.87
RetinaNet	proposed method	<b>0.9913</b>	<b>96.7</b>	<b>0.9877</b>	4.69
EfficientDet-D0	EfficientNet-B0	0.9034	89.4	0.9232	4.21
EfficientDet-D1	EfficientNet-B1	0.9243	91.5	0.9345	3.15
EfficientDet-D2	EfficientNet-B2	0.9364	94.3	0.9512	3.01
EfficientDet-D3	EfficientNet-B3	0.9621	95.4	0.9613	2.97

MAP is the average value of multiple AP categories. The larger the mAP, the better the network performance.

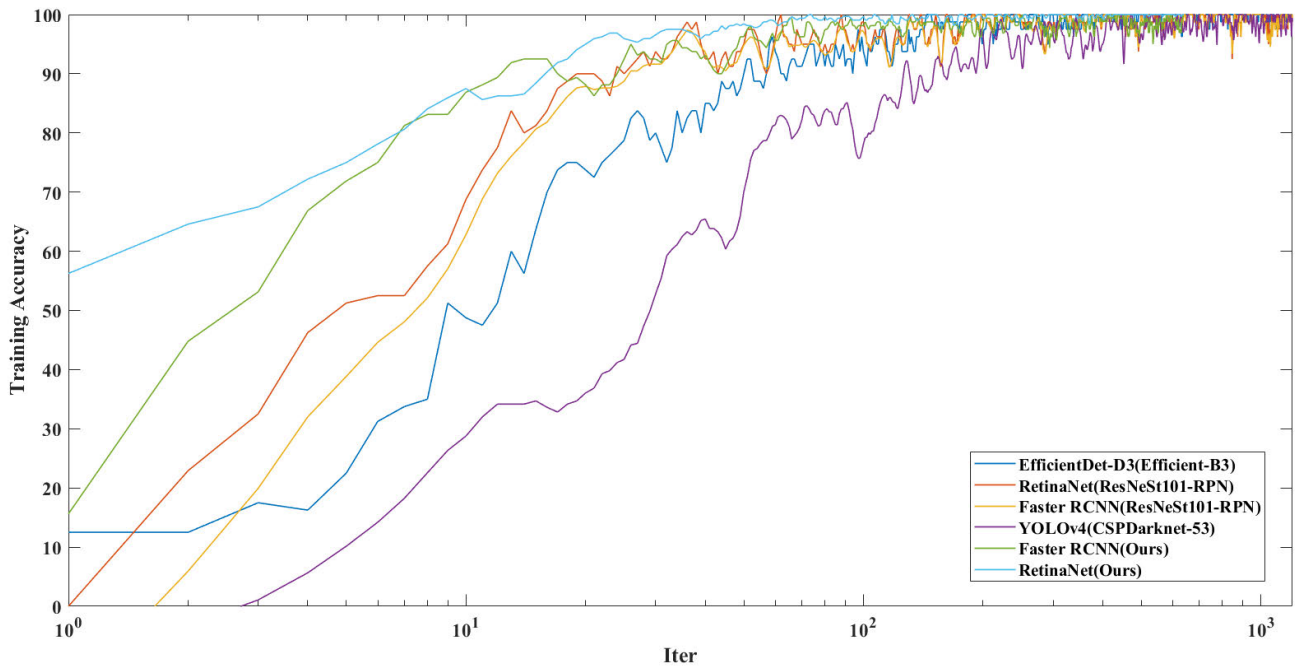
**AUC (Area Under Curve):**  $AUC = \sum_{i=2}^m ((x_i - x_{i-1}) \times (y_i + y_{i-1}) / 2)$ . AUC is the area under Receiver Operating Characteristic curve (ROC curve). The closer the AUC is to 1, the better the classifier performance is. Where  $x_i$  is the false positive rate of the abscissa of the ROC curve,  $x_i = FPR = FP / (FP + TN)$ , and  $y_i$  is the true rate of the ordinate of the ROC curve,  $y_i = TPR = TP / (TP + FN)$ .

**FPS (Frames Per Second):** FPS refers to the number of pictures processed per second. The higher the FPS, the faster the network speed.

**C. RESULTS AND EVALUATION**

A comprehensive evaluation of the proposed method is carried out on the enhanced Prates insulator data set. The performance of the proposed method is better than that of the existing methods.

In order to prove the performance of the proposed method, we compare the following defect detection algorithms. Original YOLOv3 [23], YOLOv3 by Tao, original YOLOv4 [24], Faster RCNN by Ling, Faster RCNN [25] and



**FIGURE 3.** The training accuracy of proposed network and the control group. In order to observe the rising speed of the training accuracy of all kinds of network more intuitively, the abscissa is expressed in exponential form. The proposed network rises fastest among all the training accuracy curves, and can also maintain a high degree of accuracy.

RetinaNet [26] with ResNet50, ResNet101, ResNet50-RPN, ResNet101-RPN, ResNeSt50, ResNeSt101, ResNeSt50-RPN and ResNeSt101-RPN as the backbones and four types of EfficientDet [27]. The fusion feature layer corresponding to Conv3, Conv4 and Conv5 is ResNeSt\_3d, ResNeSt\_4f and ResNeSt\_5c. The corresponding floor space dimensions are interlinked. The corresponding anchor scales of {Conv3, Conv4, Conv5} are  $\{32^2, 64^2, 128^2\}$ . In this paper, three ratios  $\{1:2, 1:1, 2:1\}$  are used. In the training, the samples with the Intersection over Union (IoU) higher than 0.7 are regarded as positive samples, and those less than 0.3 are taken as negative samples. There is parameter sharing among feature pyramid networks, which makes all levels have similar semantic information. The specific performance is evaluated in the experiment.

For the above-mentioned deep learning methods, the best hyperparameters are adjusted to obtain the best performance.

As shown in Table 1, the network with improved multi-scale RPN performs better than the original network. The ACC, mAP and AUC of Faster RCNN with proposed method as the backbone were 0.0067, 2.5% and 0.0222 higher than those of Faster RCNN with ResNeSt101-RPN. The ACC, mAP and AUC of RetinaNet with proposed method as the backbone were 0.0099, 0.8% and 0.0164 higher than those of RetinaNet with ResNeSt101-RPN. However, when proposed method is used as the backbone of our proposed network, although the accuracy is improved, the FPS decreases by 1.18. Less than 5 images are detected per second, which cannot meet the requirements for real-time detection of insulators in UAV aerial photos. The FPS of YOLOv4 can

reach 25.21. This is because YOLOv4 will first use  $1 \times 1$  convolution check to reduce feature dimension, and then use  $3 \times 3$  convolution kernel to increase dimension. In this process, the calculation number of parameters and the size of model will be greatly reduced. However, the ACC, mAP and AUC of YOLOv3 are only 0.8562, 75.6% and 0.7854, which cannot meet the requirements of real-time insulator defect detection.

We found that the break defect did not work well in all control networks, especially PGI and GGI. We list the test results of some networks which are EfficientDet-D3 (EfficientNet-B3), RetinaNet (ResNeSt101-RPN), Faster RCNN (ResNeSt101-RPN), YOLOv4 (CSPDarknet-53), Faster RCNN (Ours) and RetinaNet (Ours) on some test sets, as shown in Fig. 4. The detection speed of YOLOv4 is fast, but the accuracy is generally not high. Fast RCNN (ResNeSt101-RPN) failed to detect a small break defect on CPI, nor did it detect a piece of break defect with unclear color characteristics on GGI. RetinaNet (ResNeSt101-RPN) judged a large break defect on PGI as two break defects, and mistakenly detected the shadow of a wire on CBI as a dirt defect. EfficientDet performs well in the data set of this paper, but its speed is slow and there are many training parameters. Even the common deep learning computer card cannot train this series of networks and cannot meet the needs. For the PGI with break defect, Faster RCNN cannot detect the most part of the break defect except with the backbone of ours. The ACC detected by other networks is also low. This may be due to the massive damage of the insulator, which cannot even be identified as an insulator. For CPI with three





**FIGURE 4.** Performance of EfficientDet-D3 (EfficientNet-B3), RetinaNet (ResNeSt101-RPN), Faster RCNN (ResNeSt101-RPN), YOLOv4 (CSPDarknet-53), Faster RCNN (Ours) and RetinaNet (Ours). The first column is the insulator type, insulator position, defect category and defect location labelled by LabelImg. The text above the rectangle indicates the insulator type or defect type detected by the network. The following number indicates the accuracy of the object detected by the network.

breaks, some networks without improved RPN cannot detect the smaller defects in the middle right, such as YOLOv4. However, the improved multi-scale RPN can detect and locate such small defects.

We found that crack and dirt can be detected well in any kind of network, which may be due to the obvious characteristics of crack and dirt on the insulator. However, RetinaNet with ResNet101-RPN as the backbone and Faster RCNN with ResNet50-RPN misjudge some shadows as dirt. This may be due to the obvious contrast between shadow and dirt on the light CBI, resulting in misjudgment.

For using our proposed method as the backbone in Faster RCNN and RetinaNet, although the mAP reaches 95.8% and 96.7%, FPS is only 12.80 and 4.69. In the actual insulator defect detection, we must balance the detection accuracy and detection efficiency. We discarded the network with FPS of 4.69, leaving the network whose mAP is 95.8% and FPS is 12.8.

In order to verify the performance improvement of the improved multi-scale RPN, we compare it with the original RPN method. Four ablation experiments are designed to verify the performance of the proposed network.



- The original Faster RCNN with ResNet50 as the backbone.
- Replace ResNet50 in a) with improved ResNeSt50 as the backbone.
- The original RPN is added to the improved ResNeSt50.
- The original RPN in c) is replaced by the improved multi-scale RPN.

Table 2 shows the results of ablation experiments. Compared with the original Faster RCNN with ResNeSt50 as the backbone, the improved ResNeSt improves the mAP by 5.97% but reduces the FPS by 18.60%. The addition of improved multi-scale RPN has better accuracy in detecting insulator defects. Detection speeds have also been improved. Compared with the original RPN, the improved multi-scale RPN improves the mAP by 3.57% and FPS by 15.52%. It improves the efficiency of network detection and can meet the need of real-time detection.

**TABLE 2.** Ablation experiments of proposed network.

methods	mAP (%)	up of mAP (%)	FPS	up of FPS (%)
Faster RCNN (ResNet50)	87.1		<b>16.13</b>	
+improved ResNeSt	92.3	5.97	13.13	-18.60
+original RPN	92.5	0.21	11.08	-15.61
+improved RPN	<b>95.8</b>	3.57	12.80	15.52

## V. CONCLUSION

In this paper, a method of aerial photo insulator defect detection based on improved ResNeSt and improved multi-scale RPN is proposed. In view of the characteristics of insulator defects, the original ResNeSt is improved, and the multi-scale RPN is improved to detect smaller defects more efficiently. The experimental results show that compared with Faster RCNN with ResNeSt101-RPN as the backbone, ACC, mAP and AUC have been improved by 0.0067, 2.5% and 0.0222, and FPS have been improved by 7.57. Compared with RetinaNet with ResNeSt101-RPN as the backbone, ACC, mAP and AUC have been improved by 0.0099, 0.8% and 0.0164. In the data set of this paper, ACC, mAP and AUC have better performance than other control defect detection networks.

However, at present, there are many kinds of insulators, and only four kinds of insulators have been collected and discussed in this paper. In the follow-up work, more samples of defective insulators should be collected and added to the data set, and the network should be adjusted appropriately to adapt to more types of defect detection. In addition, the selection of a classifier for further experiments and how to improve the detection speed of the network will be further investigated as a future outlook.

## ACKNOWLEDGMENT

The authors would like to thank Prof. R. M. Prates, A. P. Marotta, Prof. E. Simas, and Prof. R. P. Ramos for their insulator data set.

## LIST OF ABBREVIATIONS

<b>HOG</b>	Histogram of Oriented Gradient
<b>LBP</b>	Local Binary Patterns
<b>RPN</b>	Region Proposal Network
<b>PCA</b>	Principal Component Analysis
<b>SVM</b>	Support Vector Machine
<b>KNN</b>	K-Nearest Neighbor
<b>ROI</b>	Region of Interest
<b>IoU</b>	Intersection over Union
<b>mIoU</b>	mean Intersection over Union
<b>FPN</b>	Feature Pyramid Networks
<b>PGI</b>	Polymeric Grey Insulator
<b>CPI</b>	Ceramic Pin Insulator
<b>GGI</b>	Glass Green Insulator
<b>CBI</b>	Ceramic Bicolor Insulator
<b>FC</b>	Fully Connected layer
<b>ACC</b>	Accuracy
<b>TP</b>	True Positives
<b>FP</b>	False Positives
<b>TN</b>	True Negatives
<b>FN</b>	False Negatives
<b>mAP</b>	mean Average Precision
<b>AUC</b>	Area Under Curve
<b>ROC</b>	Receiver Operating Characteristic
<b>FPS</b>	Frames Per Second

## REFERENCES

- X. Miao, X. Liu, J. Chen, S. Zhuang, J. Fan, and H. Jiang, "Insulator detection in aerial images for transmission line inspection using single shot multibox detector," *IEEE Access*, vol. 7, pp. 9945–9956, 2019.
- Y. Tiantian, Y. Guodong, and Y. Junzhi, "Feature fusion based insulator detection for aerial inspection," in *Proc. 36th Chin. Control Conf. (CCC)*, Jul. 2017, pp. 10972–10977.
- L. Yang, X. Jiang, Y. Hao, L. Li, H. Li, R. Li, and B. Luo, "Recognition of natural ice types on in-service glass insulators based on texture feature descriptor," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 24, no. 1, pp. 535–542, Feb. 2017.
- Z. Zhao, G. Xu, Y. Qi, N. Liu, and T. Zhang, "Multi-patch deep features for power line insulator status classification from aerial images," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 3187–3194.
- J. Katrasnik, F. Pernus, and B. Likar, "A survey of mobile robots for distribution power line inspection," *IEEE Trans. Power Del.*, vol. 25, no. 1, pp. 485–493, Jan. 2010.
- Z. Hang, C. Wu, Z. Zhang, Y. Zhu, Z. Zhang, H. Lin, Y. Sun, T. He, J. Mueller, R. Manmatha, and M. Li, "Resnest: Split-attention networks," *CoRR*, vol. abs/2004.08955, pp. 1–22, Apr. 2020.
- B. Li, J. Yan, W. Wu, Z. Zhu, and X. Hu, "High performance visual tracking with siamese region proposal network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8971–8980.
- Q. Wu and J. An, "An active contour model based on texture distribution for extracting inhomogeneous insulators from aerial images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3613–3626, Jun. 2014.
- 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 Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 4, pp. 1486–1498, Apr. 2020.
- D. Zuo, H. Hu, R. Qian, and Z. Liu, "An insulator defect detection algorithm based on computer vision," in *Proc. IEEE Int. Conf. Inf. Autom. (ICIA)*, Jul. 2017, pp. 361–365.
- M. Oberweger, P. Wohlhart, and V. Lepetit, "Training a feedback loop for hand pose estimation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3316–3324.
- L. Hu, J. Ma, and Y. Fang, "Defect recognition of insulators on catenary via multi-oriented detection and deep metric learning," in *Proc. Chin. Control Conf. (CCC)*, Jul. 2019, pp. 7522–7527.

- [13] Z. Zhao, Z. Zhen, L. Zhang, Y. Qi, Y. Kong, and K. Zhang, "Insulator detection method in inspection image based on improved faster R-CNN," *Energies*, vol. 12, no. 7, p. 1204, Mar. 2019.
- [14] L. Zenan, R. C. Qiu, Z. Jin, Y. Zhang, X. He, H. Liu, and L. Chu, "An accurate and real-time self-blast glass insulator location method based on faster R-CNN and U-net with aerial images," *CoRR*, vol. abs/1801.05143, pp. 1–8, Jan. 2018.
- [15] T. Wu, W. B. Wang, L. Yu, Y. Xie, and W. Yin, "Insulator defect detection method for lightweight YOLOV3," *Comput. Eng.*, vol. 45, no. 8, pp. 275–280, 2019.
- [16] R. M. Prates, R. Cruz, A. P. Marotta, R. P. Ramos, E. F. S. Filho, and J. S. Cardoso, "Insulator visual non-conformity detection in overhead power distribution lines using deep learning," *Comput. Electr. Eng.*, vol. 78, pp. 343–355, Sep. 2019.
- [17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [18] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1492–1500.
- [19] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2117–2125.
- [20] D. Brabandere, B. D. Neven, and L. Van Gool, "Semantic instance segmentation with a discriminative loss function," *CoRR*, vol. abs/1708.02551, pp. 1–10, Aug. 2017.
- [21] Tzutalin. *LabelImg*. Accessed: Dec. 31, 2017. [Online]. Available: <https://github.com/tzutalin/labelImg>
- [22] S. Wan, Y. Liang, and Y. Zhang, "Deep convolutional neural networks for diabetic retinopathy detection by image classification," *Comput. Electr. Eng.*, vol. 72, pp. 274–282, Nov. 2018.
- [23] R. Joseph and A. Farhadi, "Yolov3: An incremental improvement," *CoRR*, vol. abs/1804.02767, pp. 1–6, Apr. 2018.
- [24] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *CoRR*, vol. abs/2004.10934, pp. 1–17, Apr. 2020.
- [25] R. Shaoqing, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.
- [26] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
- [27] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10781–10790.



**YIFAN LIU** was born in Yichang, Hubei, in 1996. He received the B.S. degree in electrical engineering from the Hubei University of Technology, in 2018, where he is currently pursuing the M.S. degree in electrical engineering.

He is the author of two articles and four inventions. His research interests include computer vision and object detection.



**YIHUI QING** was born in Luoyang, Henan, in 1995. He received the B.S. degree in electrical engineering from the Hubei University of Technology, in 2018, where he is currently pursuing the M.S. degree in electrical engineering.

He is the author of five articles, five inventions, and one provincial key laboratory fund. His research interests include image processing and multi joint serial robot.



**CHENXI WANG** was born in Neijiang, Sichuan, in 1994. He is currently pursuing the M.S. degree in control engineering with the Hubei University of Technology. He is the author of one article. His research interests include computer vision and object detection.



**TIANZE LAN** was born in Xiaogan, Hubei, in 1993. He received the B.S. degree in electrical engineering from the Hubei University of Technology, in 2018, where he is currently pursuing the M.S. degree in electrical engineering.

He is the author of one article, one invention, and one provincial key laboratory fund. His research interests include computer vision and object detection.



**SHUQING WANG** was born in Wuhan, Hubei, in 1969. She received the B.S. degree in hydropower automation and the M.S. degree in hydraulic electro generating from Wuhan University, in 1991 and 1998, respectively, and the Ph.D. degree in system analysis and integration from the Huazhong University of Science and Technology, in 2006.

She is currently a Professor of automation with the Hubei University of Technology, the Vice President of Hubei Artificial Intelligence Society, and the Director of Central South Six Provinces Automation Society. She has published more than 60 articles and more than ten invention patents. Her research interests include artificial intelligence, intelligent control, system analysis, and integration.



**RUOTIAN YAO** received the M.S. degree in power system and automation from the Hubei University of Technology, Wuhan, China, in 2018. He is currently pursuing the Ph.D. degree with the School of Electrical Engineering and Automation, Wuhan University.

His research interests include smart grids, demand-side management, and complex networks.

• • •