IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 19 March 2023, accepted 13 April 2023, date of publication 18 April 2023, date of current version 21 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3268106

RESEARCH ARTICLE

YOLOX++ for Transmission Line Abnormal Target Detection

ZHONGQIN BI[1], LINA JING[1], CHAO SUN[1], AND MEIJING SHAN[2]
[1]College of Computer Science and Technology, Shanghai University of Electric Power, Shanghai 201306, China
[2]Institute of Information Science and Technology, East China University of Political Science and Law, Shanghai 200042, China

Corresponding author: Chao Sun (sunchao0226@163.com)

This work was supported by the Project of Shanghai Science and Technology Committee under Grant 23010501500.

ABSTRACT The detection of abnormal targets in transmission lines plays a significant role in maintaining the stability and safety of transmission systems. To achieve improved detection performance for abnormal targets, we propose a new target detector based on YOLOX, called YOLOX++. First, a multiscale cross-stage partial network (MS-CSPNet) is designed, which fuses multiscale feature information and expands the receptive field of the target through channel combination. Furthermore, depthwise and dilated convolutions are introduced in an object decoupling head to better capture the long-range dependencies of objects in feature maps. Finally, the alpha loss function (α-IoU) is introduced to optimize the localization of small objects. Experiments show that in a comparison with the YOLOX model, the YOLOX++ approach in this paper achieves 86.8% and 96.6% detection accuracies for high-voltage tower bird nest and power line insulator targets, respectively. On the PASCAL VOC dataset, the AP_{50} and AP_S are improved by 9.3% and 5.0% over those of YOLOX, respectively, showing that the YOLOX++ network possesses better robustness for small target detection.

INDEX TERMS Target detection, transmission line anomaly target, small target detection, YOLOX.

I. INTRODUCTION

In recent years, with the continuous development of social and economic development, China's power grid construction has made very significant achievements. Among them, transmission lines are equivalent to the pulse of the power system, connecting the whole set of power facilities, and are in a pivotal position in the power grid. However, due to the long-term erection of lines in the natural environment, problems such as birds nesting [1] and insulators falling off [2], [3] are prone to occur, which may lead to short circuits or open circuits in severe cases. Therefore, transmission lines must be checked regularly to maintain the stability of power transmission. Traditional power line detection methods include manual detection [4] and sensor detection [5]. Manual detection mainly relies on inspectors climbing line towers or using telescopes to inspect power lines, which is difficult and dangerous. At the same time, sensor detection requires the installation of the device in a fixed position, which is easily damaged and has low flexibility.

In recent years, the application of UAV technology [6] has alleviated the disadvantages of high costs and low efficiency of early line inspection. Inspectors now only need to operate drones to photograph transmission lines and then use image processing technology to detect the acquired images. In general, traditional image processing techniques use feature information such as texture [7], shape [8], [9], and saturation [10] to segment objects and then achieve fault detection through matching algorithms. However, the images captured by UAVs contain complex natural scenes, and the existing recognition accuracy cannot meet the requirements of industrial applications.

With the development of deep learning, image processing methods based on neural networks have gradually become mainstream. Compared with traditional methods, deep learning-based detection methods [11], [12] have better accuracy and real-time performance. However, the current detection algorithms based on deep learning are mainly aimed

The associate editor coordinating the review of this manuscript and approving it for publication was Wenbing Zhao.

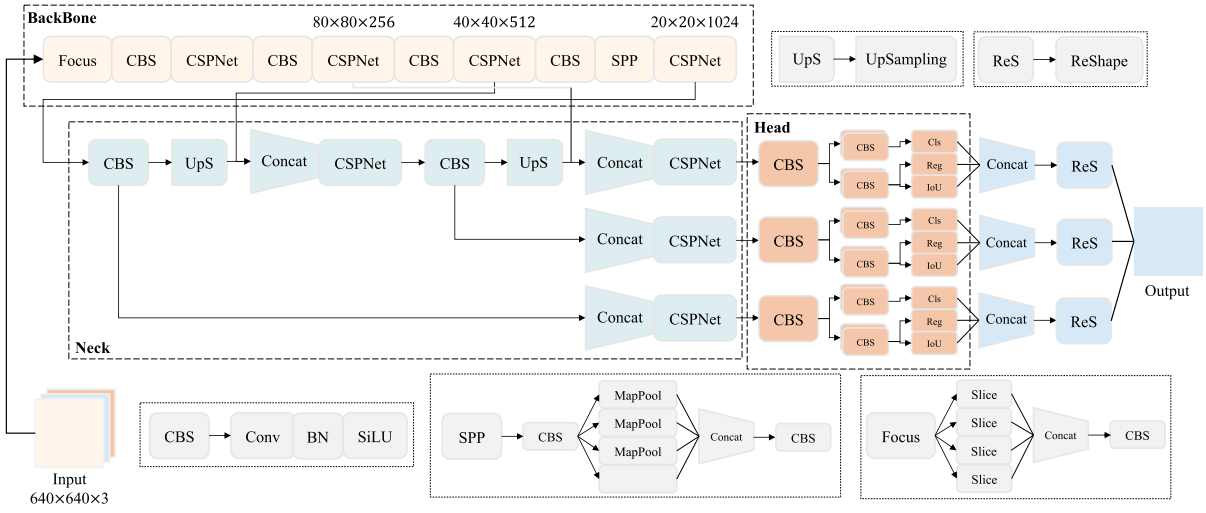


FIGURE 1. YOLOX network structure.

at detecting general objects, and the performance of the algorithm will decrease when applied to detect transmission line anomalies [13]. Compared with other targets, bird's nests [14] and insulators [15] on transmission lines have smaller physical sizes and lower contrast. Improving the accuracy of small object detection [16], [17] is also a significant challenge in current object detection tasks.

Most of the existing target detection algorithms are anchor-based detectors, and the anchor setting is not friendly to small targets. Considering the limited computing resources of embedded devices carried by UAVs, it is necessary to strictly balance detection accuracy and real-time performance. Therefore, the anchor-free [18], [19] detection algorithm YOLOX [20] is chosen as the basic network. In addition, to meet the needs of transmission line anomaly detection, the following improvements have been made to the YOLOX network:

1. We propose the multi-scale cross-stage partial network (MS-CSPNet). By using multiple feature extraction branches, obtain multi-scale feature information of the target while enhancing the feature representation of the target and further improving the detection accuracy of the target.

2. We propose a new decoupling header for the target. The target's important features and long-range dependencies are extracted by stacking and uniting convolution kernels with different functions while expanding the field-of-perception region of small targets in the feature map.

3. We reconsider the offset between target and predicted frames and use the a-IoU function to optimize the regression task of the YOLOX network to improve the accuracy of transmission line anomaly target detection.

4. Due to the scarcity of the anomaly target dataset, we constructed a bird's nest dataset and expanded the defective insulator dataset. The experimental results show that the YOLOX++ model improves the detection accuracy of bird nests and defective insulators by 9.4% and 8.9%. Meanwhile, the PASCAL VOC dataset verifies that the model has good robustness.

The rest of the paper is organized as follows: Section II introduces the related detection methods and the YOLOX network. Section III describes the specific details of the YOLOX++ network. Section IV presents the relevant experimental analysis results. Section V summarizes the results and discusses future work plans.

II. RELATED WORK

In this section, we present deep learning-based multistage and single-stage target detection algorithms, along with a detailed description of the YOLOX base network used in this paper.

A. MULTI-STAGE TARGET DETECTION ALGORITHM

The multistage detection algorithm extracts the target candidate frame and then identifies and locates the target based on the candidate frame. Early region-based convolutional neural network (RCNN) algorithms [21] utilized a selective search [22] strategy to build regions of interest (ROIs) and used SVM for object classification. After this, Fast R-CNN [23] utilizes ROI pooling to optimize regional features but increases the model complexity. Subsequently, based on Faster R-CNN [24], Wang et al. [25] extracted proposal regions through different sliding windows and detected bird's nest targets. Li et al. [26] proposed a deep learning-based framework for automatic bird nest detection-ROI mining and a faster RCNN. The detection accuracy of bird nest targets was improved by k-means clustering and the focal loss function. In addition, Zhao et al. [27] used RCNN for insulator identification and improved the detection accuracy of faults by FPN for target feature enhancement. However, the detection speed cannot meet the requirements of industrial applications because the candidate frame operations extracted

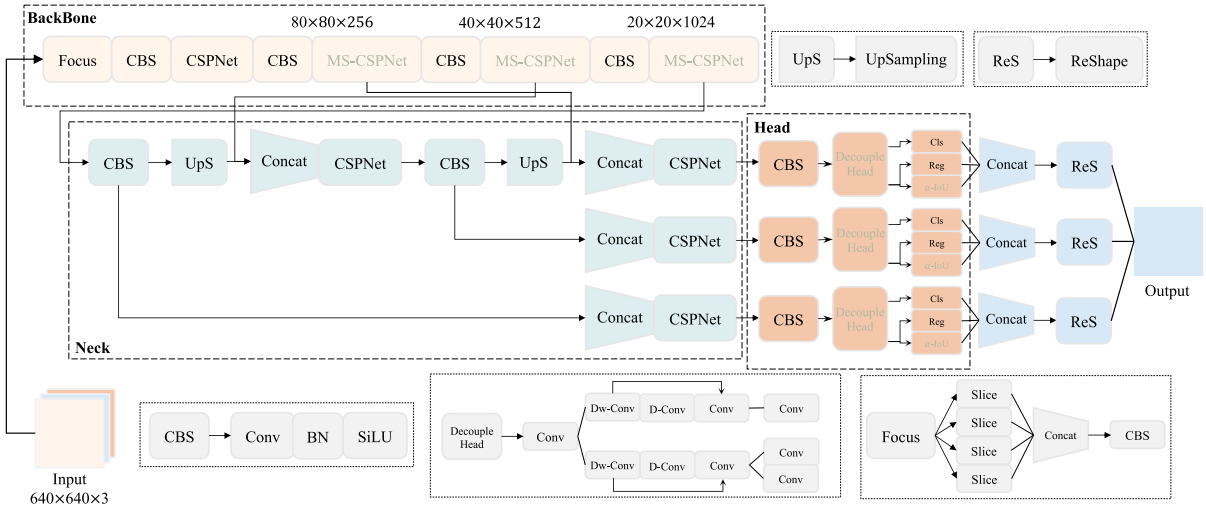


FIGURE 2. YOLOX++ network structure.

by the multistage detection algorithm often have many redundant features.

B. SINGLE-STAGE TARGET DETECTION ALGORITHM

Single-stage algorithms treat detection as a regression problem and output the detection results directly. Among them, the YOLO [28] algorithm removes the region proposal process, which greatly improves the model training speed. The SSD [29] algorithm combines the regression idea of YOLO and the anchor frame mechanism of Faster R-CNN and uses hierarchical feature maps to predict objects of different scales. To solve the problem of low detection accuracy of YOLO, YOLOv2 [30] adds an a priori frame and multiscale training strategy on its basis, which improves the overall performance of the network. Subsequently, Ying et al. [31] optimized the width and height of the prediction box and the class imbalance loss function to improve the recognition accuracy of the YOLOv3 algorithm for the bird's nest target. To detect defective insulators, Yang et al. [32] added a new feature layer to the YOLOv4 model to enrich the feature information of insulators. In addition, Liu C et al. [33] proposed a cross-stage partially dense YOLO model, which improved the detection performance of the network for insulators by improving its clustering algorithm and loss function.

C. YOLOX TARGET DETECTION ALGORITHM

To better meet the real-time requirements of abnormal target detection for transmission lines, this paper studies the YOLOX algorithm with speed advantages.

YOLOX is the YOLO series network, and its structure is shown in Fig. 1. It uses CSPDarknet [30] as a feature extraction network, enriches the feature information with an FPN [34] structure, and finally localizes and classifies objects through a decoupling head.

After the input image, the target features are extracted using the focus, superimposed CBS, and CSPNet structures. The CBS structure consists of 3×3 convolutional kernels, BN, and SiLU activation functions. CSPNet [35] uses the residual structure in ResNet [36] to segment the stacked residual blocks. In addition, an SPP module is added to the last stacking block of the backbone. SPP utilizes a parallel structure containing pooling kernels of different sizes (5×5, 9×9, 13×13), which have various sizes, to obtain the perceptual fields of the target. Finally, after multiple processing by CBS and CSPNet, three effective feature maps are generated with sizes of $80 \times 80 \times 256$, $40 \times 40 \times 512$, and $20 \times 20 \times 1024$, which are sequentially used to detect small, medium, and large objects.

After extracting effective feature information from the backbone, the feature layers at different scales are used to build a feature pyramid, which can strengthen the extraction of target feature information. The specific details include upsampling the deep feature map, merging it with the midlevel feature channel, and performing an upsampling operation to match the shallow feature channel. Through a top-down information flow, the semantic target information can be preserved in shallow feature maps. To enhance its target feature expression ability, YOLOX integrates PANet [37] into an FPN to construct a bottom-up feature transfer path to further enrich the target feature information.

Compared with other detectors, YOLOX uses a more convergent decoupling head in the detection head component [38], [39]. The decoupling head divides the input features and processes the target's classification task and localization task in parallel. Finally, the target detection task is completed using the prediction information obtained by different branches.

III. THE PROPOSED YOLOX++ MODEL

To better detect small targets on transmission lines, this paper proposes the YOLOX++ model, which enhances the multiscale sensory fields of targets through an MS-CSPNet

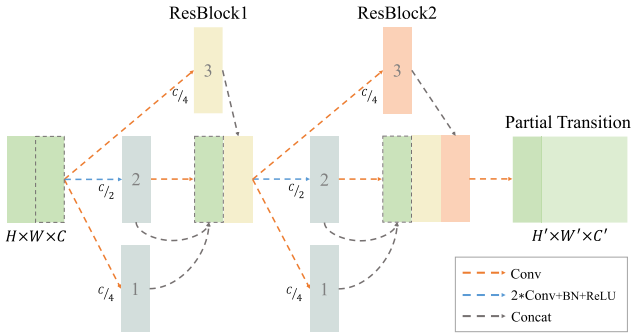


FIGURE 3. Multi-scale cross stage partial ResNet (MS-CSPNet).

while constructing a new target decoupling head for small targets and finally improving the target localization effect by optimizing the loss function. The network structure of YOLOX++ is shown in Fig. 2.

A. MULTI-SCALE CROSS STAGE PARTIAL RESNET

The backbone network of YOLOX is CSPDarknet, which extracts target features by stacking multiple CSPNets. To explore the feature information of the target, this paper proposes a multi-scale cross-stage partial ResNet (MS-CSPNet), a novel network structure that learns the target features at different scales. MS-CSPNet splits the input into multiple branches and employs stacked convolutional kernels to each branch to enhance the feature information. Moreover, MS-CSPNet utilizes residual networks to increase the network depth and improve detection accuracy. Notably, MS-CSPNet adopts cross-stage channel connections to mitigate the vanishing gradient problem caused by the depth increment.

The primary purpose of designing the MS-CSPNet is to allow the architecture to achieve rich gradient combinations while enhancing the perceptual fields of small targets. As shown in Fig. 3, the input of $H \times W \times C$ is truncated into three parts. Path 1 uses a 3×3 convolutional kernel to extract target features and reduces the input channels to C/4. Path 2 enhances the target representation by stacking 3×3 convolutional kernels, BN layers and ReLU activation functions and splits the output into two parts: one part of the target information awaits merging, and the other part continues with feature extraction. Path 3 uses residual blocks to mitigate gradient variations and provide the output. Finally, the channel information in different paths is merged, and the effective feature maps at different scales are output.

During the feature information transfer process, stacked 3×3 convolutional kernels are used instead of 5×5 and 7×7 convolutional kernels to further enhance the feature extraction effect for small targets. Furthermore, retaining the original cross-stage strategy can alleviate the network overhead caused by explicit feature maps. To avoid excessive feature redundancy, we use the MS-CSPNet only in the effective feature map layer and remove the final SPP structure to enhance the extraction speed for target features.

Overall, through its splitting and merging strategy, the proposed MS-CSPNet retains the advantage of feature reusability and enhances the perceptual field areas of small targets in the feature map. It also prevents the presence of too much repetitive gradient information by truncating the gradient flow and finally uses a simple convolution operation to integrate the extracted feature information.

B. TARGET DECOUPLING HEAD

After obtaining feature information from the feature extraction network, the feature layers at different scales are passed into the YOLOX detection head to complete the classification and regression tasks. In this paper, we follow two branches of classification and regression, using depthwise convolution to further extract the local information of the target, dilation convolution to improve the perceptual field region of the target in the feature layers, and subsequently joint input feature layers to capture the remote dependence of the target. Finally, the prediction results of each feature layer are obtained by convolutional layers. By improving the detection head of YOLOX and making full use of the feature information of the target, the detection effect of the network can be improved while reducing the number of parameters.

As shown in Fig. 4, the effective feature maps of different feature channels enter the detection head. First, a 1×1 convolutional layer is used to reduce the channel dimension. The feature channels are then processed through two parallel depthwise convolutional layers, a dilated convolutional layer and a 1×1 convolutional layer. For each feature layer, obtain feature point's regression parameters and classification category. Finally, the small object detection task is accomplished by stacking multiple branches.

In the decoupling head, the output is fused with the local information processed by the 1×1 convolutional layer to better capture the long-range target information. The calculation process is as follows:

$$F' = Conv_{1\times1}\left(DConv\left(DWConv\left(F\right)\right)\right) \tag{1}$$

$$Output = F' \otimes F \tag{2}$$

Here, $F \in R^{(C\times H\times W)}$ is the input feature, and $\otimes$ refers to the product of elements. The detection head of YOLOX++ can retain more target feature points while reducing the number of network parameters and strengthening the classification and positioning of small targets.

C. LOSS FUNCTION

In the anchor-free detector, YOLOX reduces the number of predicted bounding boxes from 3 to 1 for each location and makes them directly predict the offset value. This improvement reduces the number of parameters and the computational effort of the detector but also obtains better performance. However, for anomalous targets on transmission lines, which are more challenging to locate than ordinary objects

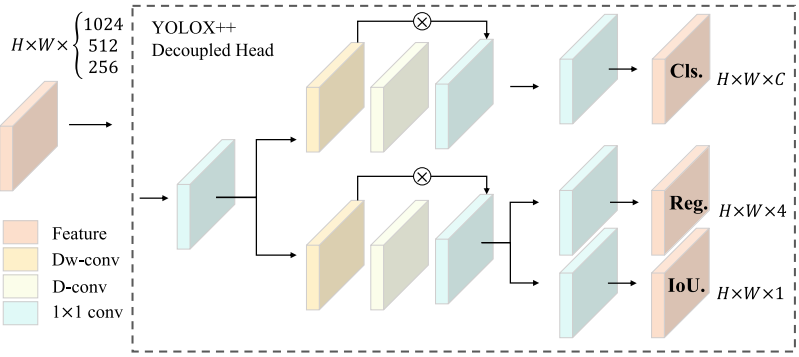


FIGURE 4. Target decoupling head.

due to their lower resolution, a better evaluation of the offset of the prediction frame from the target frame is needed for detection.

In YOLOX, the loss function contains three components: classification loss, bounding box regression loss, and confidence loss. Among them, the binary cross-entropy loss function is used to calculate the loss of category probability and target confidence score, the IoU function is used as the loss of bounding box regression, and the total loss function is as follows:

$$L_{YOLOX} = L_{cls} + L_{reg} + L_{obj} \tag{3}$$

To improve the target localization accuracy, this paper uses the α-IoU loss [40] function for regression calculation and combines the cross-entropy loss function to train the detection branch together. The α-IoU function is calculated as follows:

$$L_{\alpha-IoU} = \frac{1 - IoU^{\alpha}}{\alpha}, \quad \alpha > 0 \tag{4}$$

Among them, α is a controllable parameter.

When $\alpha \geq 1$, the power penalty/regularization term is introduced in the formula, which can be extended to a general expression:

$$L_{\alpha-IoU} = 1 - IoU^{\alpha 1} + \rho^{\alpha 2}(B, B^{gt}) \tag{5}$$

Here, $\alpha 1 > 0$, $\alpha 2 > 0$, and $\rho^{\alpha 2}(B, B^{gt})$ denotes any penalty term computed based on B and B^{gt}. This simple extension enables the straightforward generalization of existing IoU based losses to their α-IoU versions.

Compared with the IoU loss, the α-IoU loss can yield improved regression accuracy for small targets by adaptively enhancing the weights of the loss and gradient through the deployment of α and then improve the detection accuracy of the YOLOX++ model for transmission line anomaly targets.

IV. ANALYSIS OF EXPERIMENTAL RESULTS

A. IMPLEMENTATION DETAILS

The computer processing model used in the experiment includes an Intel i9-9900K CPU, an NVIDIA GeForce 2080 GPU, 16 GB of memory, and the Ubuntu 18.04 operating system, and the PyTorch deep learning framework is used to train the network. The experimental parameters used for training the network are shown in Table 1.

TABLE 1. Experimental parameter settings.

Parameter	Value
Input image size	640×640
Data batches	4
Number of iterations	120
Mosaic data augmentation	True
Optimizer	SGD
Momentum	0.937
Weight decay	5e-4

To evaluate the effectiveness of the YOLOX++ model in terms of abnormal target detection for transmission lines, the detection accuracy (AP), F1 score, and detection speed (FPS) are selected as evaluation indicators:

$$\begin{cases} AP = \int_0^1 P(R)dR \\ F1 = 2\dfrac{PR}{P+R} \end{cases} \tag{6}$$

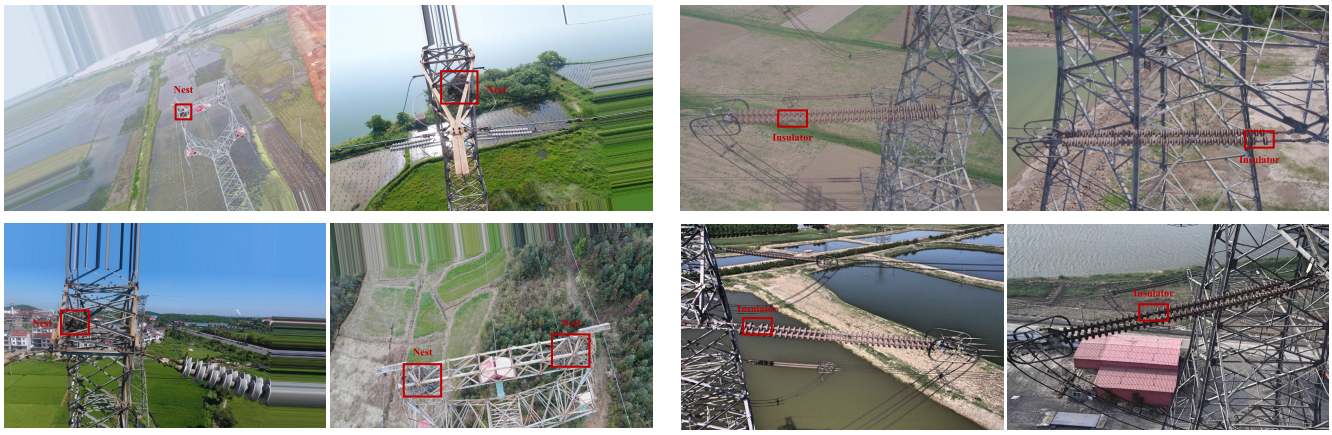
where P is the accuracy rate of target detection, and R is the recall rate:

$$\begin{cases} P = \dfrac{TP}{TP+FP} \\ R = \dfrac{TP}{TP+FN} \end{cases} \tag{7}$$

The TP term represents the number of samples where the detected target category is consistent with the actual target category. The FP term is the number of examples where the detected target category is inconsistent with the true target category. The FN term is the number of samples where the actual target exists but has not been detected by the network.

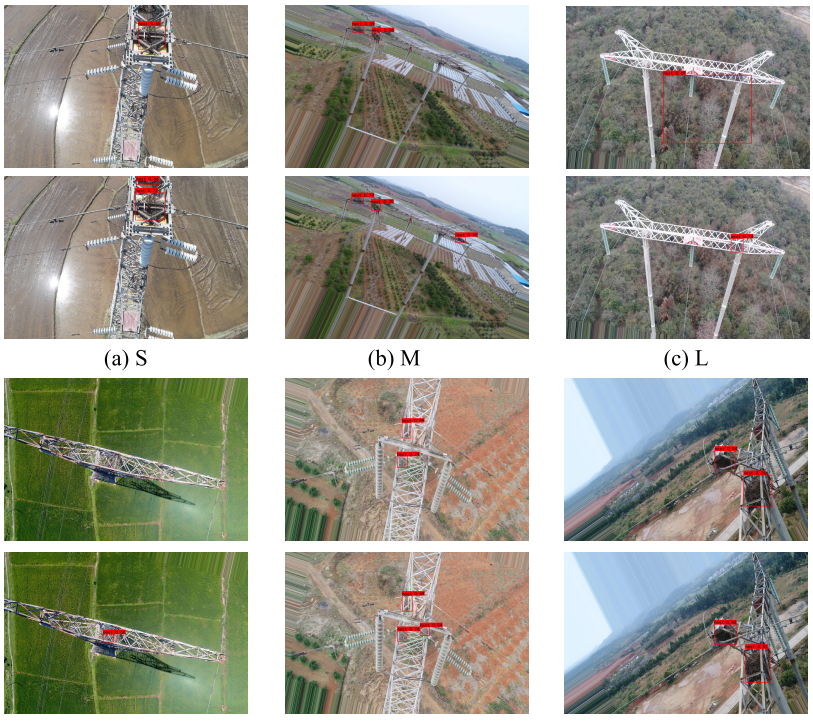
B. EXPERIMENTAL DATASET

We perform all experiments on different data benchmarks, i.e., a high-voltage tower bird nest dataset, a Chinese power line insulator dataset (CPLID) [41], and PASCAL VOC [42].



(a) High Voltage Tower Bird-nest (b) Power Line Insulator

FIGURE 5. Illustrations of a high-voltage tower bird nest and a power line insulator.



(a) S (b) M (c) L

(d) X (e) Tiny (f) nano

FIGURE 6. Visualization of the YOLOX and YOLOX++ model detection results (the top shows the YOLOX model, and the bottom shows the YOLOX++ model).

The high-voltage tower bird nest dataset is self-built, and the detection targets are high-voltage tower bird nests observed under different weather and lighting conditions. To enrich the image dataset, this paper uses image enhancement techniques (cropping, flipping, rotation, etc.) for expansion purposes. There are 2,864 images for training and 716 images for testing network performance.

While the CPLID dataset provides 600 images of insulators and 248 images of defective insulators, we perform the same data augmentation operations we end up with 2800 images for training and 700 images as the test set. In the high-voltage tower bird's nest and transmission line insulator datasets, all images are uniformly cropped to 1368 × 912. Additionally, we annotate the expanded parts of both datasets to make them conform to the standard format of PASCAL VOC. Relevant examples from the dataset are shown in Fig. 5.

C. EXPERIMENTAL RESULTS ON THE BIRD NEST DATASET

We first conduct an experimental study on the bird nest data benchmark, where different versions of the YOLOX model are trained under the same experimental conditions and analysed in comparison with the mainstream detection models.

TABLE 2. Comparison among different versions of the YOLOX and YOLOX++ models.

Models	F1	Parameters	AP(%)	AP_{75}	AP_S	AP_M	AP_L
YOLOX-S	0.76	8.9 M	78.8	19.1	15.3	31.2	46.4
YOLOX++S	0.86	8.7 M	**85.9** (+7.1)	31.2	19.6	37.5	54.6
YOLOX-M	0.80	25.3 M	81.4	22.7	17.1	34.1	49.1
YOLOX++M	0.87	24.7 M	**86.3** (+4.9)	29.0	20.6	37.4	54.0
YOLOX-L	0.76	54.2 M	77.5	23.7	17.1	33.9	48.7
YOLOX++L	0.85	53.1 M	**85.2** (+7.7)	30.6	20.4	37.3	54.9
YOLOX-X	0.73	99.1 M	77.4	19.0	15.2	30.9	44.8
YOLOX++X	0.85	97.4 M	**86.8** (+9.4)	30.8	21.0	37.3	55.3
YOLOX-Tiny	0.81	5.1 M	83.8	24.8	19.0	33.8	51.7
YOLOX++Tiny	0.85	4.9 M	**84.9** (+1.1)	27.7	18.9	35.2	53.9
YOLOX-nano	0.69	0.9 M	75.2	16.1	13.5	28.2	44.2
YOLOX++nano	0.81	1.1 M	**81.7** (+6.5)	17.8	14.3	31.6	47.8

TABLE 3. Comparative analysis of different models.

Models	Backbone	Size	FPS	Latency	AP(%)
YOLOv3 [43]	Darknet-53	608	22.5	44.4ms	69.5
YOLOv3 [43]	Darknet-53	800	15.4	64.9ms	71.9
YOLOv4 [44]	Darknet-53	608	32.5	30.7ms	71.7
EfficientDet-D0 [45]	Efficient-B0	512	50.4	19.8ms	60.4
EfficientDet-D1 [45]	Efficient-B0	640	39.6	25.2ms	65.3
PP-YOLO [46]	ResNet50-vd-dcn	512	49.5	20.2ms	72.1
PP-YOLO [46]	ResNet50-vd-dcn	608	43.2	23.1ms	75.5
PP-YOLOv2 [47]	ResNet50-vd-dcn	640	35.6	28.0ms	75.6
PP-YOLOv2 [47]	ResNet101-vd-dcn	640	29.8	33.5ms	77.5
YOLOv5-S	CSPDarknet-53	640	55.4	18.0ms	73.2
YOLOv5-M	CSPDarknet-53	640	49.3	20.2ms	74.1
YOLOv5-L	CSPDarknet-53	640	28.9	34.6ms	74.9
YOLOv5-X	CSPDarknet-53	640	18.7	53.4ms	75.3
YOLOX++S	Modified CSPDarknet-53	640	68.7	14.5ms	**85.9**
YOLOX++M	Modified CSPDarknet-53	640	41.1	24.3ms	**86.3**
YOLOX++L	Modified CSPDarknet-53	640	25.0	40.0ms	**85.2**
YOLOX++X	Modified CSPDarknet-53	640	16.2	61.7ms	**86.8**
YOLOX++Tiny	Modified CSPDarknet-53	640	69.1	14.4ms	**84.9**
YOLOX++nano	Modified CSPDarknet-53	640	55.8	17.9ms	**81.7**

TABLE 4. Results of ablation experiments.

Method	AP(%)	Parameters	Latency	FPS
YOLOX-S	78.8	9.0 M	13.7ms	72.5
+MS-CSPNet	**81.7**	9.4 M	14.4ms	69.4
+Decoupling head	**82.1**	8.2 M	14.1ms	70.5
+α-IoU	**79.6**	9.0 M	13.9ms	71.9
YOLOX++S	**85.9**	8.7 M	14.5ms	68.7

TABLE 5. Overall performance comparison analysis.

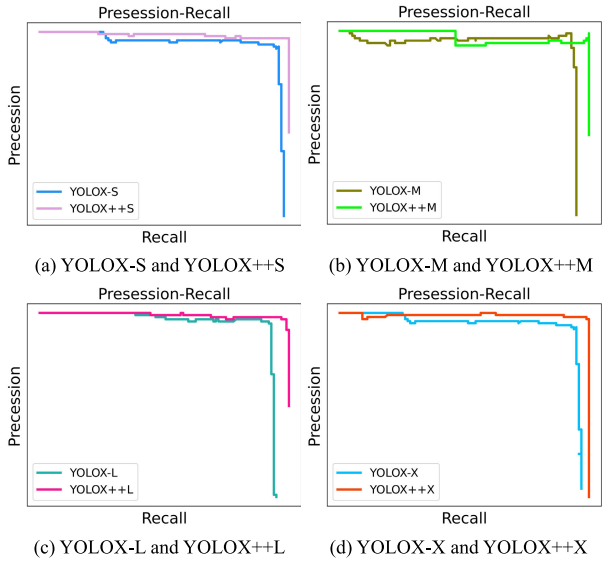
Models	AP(%)	Parameters	Latency	FPS
YOLOX-S	90.0	8.9 M	13.8ms	72.4
YOLOX++S	**96.6**	8.7 M	14.6ms	68.2
YOLOX-M	86.6	25.3 M	23.8ms	42.0
YOLOX++M	**95.5**	24.7 M	24.7ms	40.4
YOLOX-L	90.0	54.2 M	38.7ms	25.8
YOLOX++L	**95.7**	53.1 M	39.3ms	25.4
YOLOX-X	89.5	99.1 M	58.8ms	17.0
YOLOX++X	**95.7**	97.4 M	60.9ms	16.4

1) COMPARATIVE PERFORMANCE ANALYSIS

Table 2 shows the detection results of the YOLOX models with different depths and widths (S, M, L, X) for bird nests in high-voltage towers. In terms of the AP and F1 metrics, the different YOLOX++ detection models are superior to the YOLOX network. Among them, the F1 score of the YOLOX-X model with the most parameters is improved by 12%, the detection accuracy is improved by 9.4%, and the

TABLE 6. Comparison of the detection results achieved on the PASCAL VOC dataset.

Model	FPS	Latency	AP(%)	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
YOLOX-S	70.9	14.0ms	37.2	65.0	38.1	25.2	31.7	39.9
YOLOX++S	67.1	14.8ms	46.6	**74.3** (+9.3)	49.2	**30.2** (+5.0)	35.7	51.1



(a) YOLOX-S and YOLOX++S (b) YOLOX-M and YOLOX++M

(c) YOLOX-L and YOLOX++L (d) YOLOX-X and YOLOX++X

FIGURE 7. The PR curves of YOLOX and YOLOX++ models.

AP reaches 86.8%. We further reduce the model to YOLOX-Tiny and obtain a 1.1% AP improvement. The YOLOX++ nano model for mobile devices has an accuracy of 81.7% for target detection, with only 1.1 M parameters. As shown in Table 2, the YOLOX++ model possesses better detection performance than YOLOX.

Based on the test set of the high-voltage tower bird's nest (shown in Fig. 6), we demonstrate that the YOLOX++ detection algorithm can locate the target more accurately than the YOLOX baseline while improving the detection accuracy of the bird's nest target.

2) COMPARISON WITH OTHER DETECTION MODELS

To further verify the effectiveness of the YOLOX++ network, we compare it with mainstream detectors, and the results are shown in Table 3.

According to Table 3, we find that the version of YOLOv3 with Darknet-53 as the backbone has a detection accuracy of 71.9% for bird nests, and it takes 64.9 ms to detect each image. The PP-YOLO model with ResNet50 as the backbone achieves a detection accuracy of 75.5% at a detection speed of 43.2 FPS. Under the same input image, the detection accuracy of PP-YOLOv2 with a deeper backbone reaches 77.5%. In contrast, the modified YOLOX++ network based on the YOLOX baseline achieves a better balance between accuracy and speed.

3) ABLATION EXPERIMENTS

In this section, we incrementally present the effectiveness of each module, and the experimental results are shown in Table 4.

The AP of the detector with YOLOX-S, as the baseline for target detection, is 78.8%, and its FPS is 72.5. When using the MS-CSPNet to extract the multiscale features of small targets, a 2.9% accuracy improvement is obtained, but the FPS drops to 69.4 due to the increased number of network parameters. By replacing the detection head and capturing the contextual information of the target in the feature map, the detection accuracy can be improved to 82.1% while reducing the number of network parameters to 8.2 M. Finally, optimizing the regression loss of the network for small targets can improve the detection accuracy of the YOLOX++S network for bird nests in high-voltage towers to some extent.

As show in Table 4, YOLOX++S improves the detection accuracy to 85.9%, and the detection speed is 68.7 FPS, indicating that the improved network in this paper can effectively improve the detection performance of small objects in the transmission line.

D. EXPERIMENTAL RESULTS ON THE INSULATOR DATASET

The improved YOLOX++ network achieves better detection accuracy and speed on the high-voltage tower bird's nest dataset with occlusion problems. To verify the effectiveness of YOLOX++ for other object detection, we conducted related experiments on the CPLID data benchmark, and the results are shown in Table 5.

Defective insulators have smaller physical dimensions and less feature information than bird's nest targets. However, the YOLOX-S algorithm achieves an insulator detection accuracy of 90% due to the absence of obscurants. Meanwhile, our YOLOX++S model improves the detection accuracy by 6.6% and reaches the highest level among the four baseline models. Although the method in this paper can effectively achieve improved detection accuracy for small targets in real-time industrial applications, the detection speed is still a factor to be considered, which is also an issue that needs further improvement in this paper.

In addition, as seen from the PR curve in Fig. 7, the improved YOLOX++ network based on the YOLOX model improves the detection accuracy and recall of objects to a new level. To better verify the detection ability of the YOLOX++ network for defective insulators, we show the visualization results in Fig. 8. Compared with the YOLOX network, YOLOX++ in this paper can detect smaller insulator targets and simultaneously improve the target detection accuracy.

E. ROBUSTNESS TO SMALL TARGETS

Considering that the targets in both the high-voltage tower bird nest dataset and the CPLID dataset are small targets belonging to a single category, to better illustrate the

(a) S (b) M (c) L (d) X

FIGURE 8. Visualizations of the YOLOX and YOLOX++ model detection results (the top part contains the YOLOX results, and the bottom part contains the YOLOX++ model results).

TABLE 7. Detection results obtained for different data categories.

Category	YOLOX-S				YOLOX++S			
	F1	R(%)	P(%)	AP(%)	F1	R(%)	P(%)	AP(%)
aero plane	0.79	70.2	89.2	83.5	083	76.6	90.0	86.1
bicycle	0.68	63.5	74.1	76.1	0.85	81.0	89.5	**89.0**
bird	0.54	42.5	75.5	59.0	0.64	52.9	82.1	**68.6**
boat	0.48	35.4	73.9	49.3	0.59	45.8	81.5	**64.8**
bottle	0.44	31.9	69.8	43.1	0.52	37.9	83.0	**53.6**
bus	0.66	61.1	71.0	61.7	0.73	66.7	80.0	**74.6**
car	0.79	70.5	89.1	80.9	0.82	74.3	90.9	81.9
cat	0.64	59.7	69.7	60.5	0.69	63.6	74.2	**74.9**
chair	0.50	37.5	73.1	53.7	0.56	42.8	82.3	**62.5**
cow	0.70	62.5	79.6	70.4	0.73	75.0	71.2	74.4
dining table	0.70	75.0	65.2	72.7	0.83	72.5	72.5	78.9
dog	0.46	32.7	80.0	58.7	0.64	53.6	79.7	**73.3**
horse	0.78	71.4	87.0	80.0	0.86	80.4	91.8	**90.9**
motorbike	0.62	49.4	81.5	67.8	0.72	61.8	85.9	**78.0**
person	0.76	67.3	96.2	81.2	0.80	72.4	88.9	84.1
potted plant	0.42	29.8	69.4	39.7	0.46	34.5	67.4	**46.1**
sheep	0.48	37.0	66.7	49.1	0.56	48.2	66.7	**63.8**
sofa	0.45	32.7	73.9	60.0	0.66	55.8	80.6	**76.2**
train	0.71	61.5	85.1	74.0	0.81	70.8	93.9	**81.2**
tv monitor	0.74	66.1	84.8	78.0	0.78	69.5	89.1	83.4

robustness of YOLOX++ to small targets, we conduct a study on the PASCAL VOC dataset, and the related results are shown in Table 6.

The AP_{50} achieved by the YOLOX-S network on the VOC dataset is 65%, and the detection accuracy for small targets less than 32 pixels in size is only 25.2%. The network is improved by the method in this paper. The AP_{50} obtained by YOLOX++S on this dataset is 74.3%, and the AP_S is improved by 5.0%. In addition, according to the detection results for each target category in Table 7, the

YOLOX++ model can improve the detection accuracies obtained for small targets such as "bird," "boat," and "bottle" by 9.6%, 15.5%, and 10.5%, respectively, indicating that the YOLOX++ model in this paper has better robustness to small object detection.

V. CONCLUSION

In this work, we take bird's nests and defective insulator targets on transmission lines as research objects and combine them with YOLOX target detection algorithms to solve the problem of low accuracy and poor robustness of anomaly target detection in power scenarios.

In this paper, we propose a YOLOX++ detector based on MS-CSPNet. First, a new feature extraction network is used to obtain more feature information about the target and improve the detection accuracy. In addition, in the detection head, the classification and localization of small objects are enhanced by depth and dilation convolution and joint input features. Finally, the α-IoU loss function is used to improve the regression accuracy of the target. The experimental results show that the YOLOX++ model significantly improved the detection accuracy of bird nests and defective insulators in high-voltage towers compared with mainstream detectors such as YOLOX, with mAP of 86.8% and 96.6%, respectively. In addition, the experimental results of PASCAL VOC show that the proposed YOLOX++ model has better application value for small targets and can effectively solve the detection problems caused by small size and foreign object occlusion of abnormal targets on transmission lines.

Although the YOLOX++ network has achieved good detection results for abnormal targets on power lines, real-time detection still needs to be considered. In future research, we will continue to propose a lightweight feature extraction network and feature enhancement network to significantly improve the detection's real-time performance based on improving the target detection accuracy. In addition, only two transmission line anomaly targets are investigated in this paper. More anomaly targets will be tried in the next step to continuously improve our method and enhance the detection capability of the YOLOX model for transmission line anomaly targets.

REFERENCES

[1] W. Yu, X. Li, and T. Yu, "Iterative recognition of bird's nest in aerial photograph of high voltage transmission tower," *Recent Adv. Comput. Sci. Commun. Formerly, Recent Patents Comput. Sci.*, vol. 15, no. 6, pp. 851–858, 2022.

[2] K. Zhang, S. Qian, J. Zhou, C. Xie, J. Du, and T. Yin, "ARFNET: Adaptive receptive field network for detecting insulator self-explosion defects," *Signal, Image Video Process.*, vol. 16, no. 9, pp. 1–9, 2022.

[3] Q. Wen, Z. Luo, R. Chen, Y. Yang, and G. Li, "Deep learning approaches on defect detection in high resolution aerial images of insulators," *Sensors*, vol. 21, no. 4, p. 1033, 2021.

[4] L. Matikainen, M. Lehtomäki, E. Ahokas, J. Hyyppä, M. Karjalainen, A. Jaakkola, A. Kukko, and T. Heinonen, "Remote sensing methods for power line corridor surveys," *ISPRS J. Photogramm. Remote Sens.*, vol. 119, pp. 10–31, Sep. 2016.

[5] L. Erhan, M. Ndubuaku, M. D. Mauro, W. Song, M. Chen, G. Fortino, O. Bagdasar, and A. Liotta, "Smart anomaly detection in sensor systems: A multi-perspective review," *Inf. Fusion*, vol. 67, pp. 64–79, Mar. 2021.

[6] R. Bhola, N. H. Krishna, K. Ramesh, J. Senthilnath, and G. Anand, "Detection of the power lines in UAV remote sensed images using spectral–spatial methods," *J. Environ. Manag.*, vol. 206, pp. 1233–1242, Jan. 2018.

[7] W. Wang, Y. Wang, J. Han, and Y. Liu, "Recognition and drop-off detection of insulator based on aerial image," in *Proc. 9th Int. Symp. Comput. Intell. Design (ISCID)*, vol. 1, 2016, pp. 162–167.

[8] Y. Zhai, R. Chen, Q. Yang, X. Li, and Z. Zhao, "Insulator fault detection based on spatial morphological features of aerial images," *IEEE Access*, vol. 6, pp. 35316–35326, 2018.

[9] S. Liao and J. An, "A robust insulator detection algorithm based on local features and spatial orders for aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 963–967, May 2015.

[10] X. Zhang, J. An, and F. Chen, "A simple method of tempered glass insulator recognition from airborne image," in *Proc. Int. Conf. Optoelectronics Image Process.*, vol. 1, Nov. 2010, pp. 127–130.

[11] B. Chen and X. Miao, "Distribution line pole detection and counting based on YOLO using UAV inspection line video," *J. Elect. Eng. Technol.*, vol. 15, no. 1, pp. 441–448, 2020.

[12] 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.

[13] Y. Cao, "VisDrone-DET2021: The vision meets drone object detection challenge results," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2021, pp. 2847–2854.

[14] Q. Li, F. Zhao, Z. Xu, J. Wang, K. Liu, and L. Qin, "Insulator and damage detection and location based on YOLOV5," in *Proc. Int. Conf. Power Energy Syst. Appl. (ICoPESA)*, 2022, pp. 17–24.

[15] Z. Ge, H. Li, R. Yang, H. Liu, S. Pei, Z. H. Jia, and Z. Ma, "Bird's nest detection algorithm for transmission lines based on deep learning," in *Proc. 3rd Int. Conf. Comput. Vis., Image Deep Learn. Int. Conf. Comput. Eng. Appl. (CVIDL ICCEA)*, 2022, pp. 417–420.

[16] S. S. Rawat, S. Alghamdi, G. Kumar, Y. Alotaibi, O. I. Khalaf, and L. P. Verma, "Infrared small target detection based on partial sum minimization and total variation," *Mathematics*, vol. 10, no. 4, p. 671, 2022.

[17] I. Singh and G. Munjal, "Improved Yolov5 for small target detection in aerial images," *SSRN 4049533*, to be published.

[18] H. Law and J. Deng, "Cornernet: Detecting objects as paired keypoints," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 734–750.

[19] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," 2019, *arXiv:1904.07850*.

[20] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021, *arXiv:2107.08430*.

[21] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[22] J. R. R. Uijlings, "Selective search for object recognition," *Int. J. Comput. Vis.*, vol. 104, no. 2, pp. 154–171, Feb. 2013.

[23] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jun. 2015, pp. 1440–1448.

[24] S. Ren, 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.*, vol. 28, 2015, pp. 1–9.

[25] J. Wang, H. Luo, P. Yu, L. Zheng, and F. Hu, "Bird's nest detection in multi-scale of high-voltage tower based on faster R-CNN," *J. Beijing Jiaotong Univ.*, vol. 43, no. 5, pp. 37–43, 2019.

[26] J. Li, "Deep learning-based bird's nest detection on transmission lines using UAV imagery," *Appl. Sci.*, vol. 10, no. 18, p. 6147, 2020.

[27] W. Zhao, M. Xu, X. Cheng, and Z. Zhao, "An insulator in transmission lines recognition and fault detection model based on improved faster RCNN," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 5016408.

[28] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 779–788.

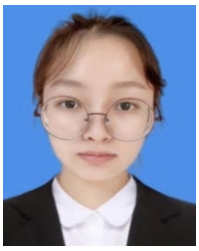
[29] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 21–37.

[30] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2017, pp. 7263–7271.

[31] Z. Yingchun, S. Si-Yu, L. Shuai, L. Zhi-Yong, X. Yong-Liang, and H. Hui-Qing, "Recognition of bird's nest on transmission tower in aerial image of high-voltage power line by YOLOV3 algorithm," *J. Guangdong Univ. Technol.*, vol. 37, no. 3, pp. 42–48, 2020.

[32] Q. Yang, X. Rong, R. Guo, H. Zhao, and Y. Zhao, "Real-time detection of insulators and drop fuses based on improved YOLOv4," *Sci. Program.*, vol. 2022, Mar. 2022, Art. no. 5755265.
[33] 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," *Appl. Sci.*, vol. 11, no. 10, p. 4647, 2021.
[34] 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.
[35] C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2020, pp. 390–391.
[36] S. Targ, D. Almeida, and K. Lyman, "Resnet in Resnet: Generalizing residual architectures," 2016, *arXiv:1603.08029*.
[37] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8759–8768.
[38] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2980–2988.
[39] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully convolutional one-stage object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2019, pp. 9627–9636.
[40] J. He, S. Erfani, X. Ma, J. Bailey, Y. Chi, and X.-S. Hua, "Alpha-IoU: A family of power intersection over union losses for bounding box regression," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 1–13.
[41] A. Raimundo, "Insulator data set-Chinese power line insulator dataset (CPLID)," *IEEE Dataport*, 2020.
[42] M. Everingham, A. Zisserman, C. K. Williams, L. Gool, M. Allan, C. M. Bishop, O. Chapelle, N. Dalal, T. Deselaers, and G. Dorkó, "The 2005 Pascal visual object classes challenge," in *Proc. Mach. Learn. Challenges Workshop*. Cham, Switzerland: Springer, 2005, pp. 117–176.
[43] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
[44] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
[45] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Nov. 2020, pp. 10781–10790.
[46] X. Long, K. Deng, G. Wang, Y. Zhang, Q. Dang, Y. Gao, H. Shen, J. Ren, S. Han, E. Ding, and S. Wen, "PP-YOLO: An effective and efficient implementation of object detector," 2020, *arXiv:2007.12099*.
[47] X. Huang, X. Wang, W. Lv, X. Bai, X. Long, K. Deng, Q. Dang, S. Han, Q. Liu, X. Hu, D. Yu, Y. Ma, and O. Yoshie, "PP-YOLOv2: A practical object detector," 2021, *arXiv:2104.10419*.

ZHONGQIN BI received the Ph.D. degree in system analysis and integration from East China Normal University, China, in 2009. He is currently a Professor with the College of Computer Science and Technology, Shanghai University of Electric Power. He has published more than 30 papers in refereed journals and conference proceedings. His research interests include cloud computing, data process, and quality control in smart grid.



LINA JING was born in Jiangsu, China, in 1997. He is currently pursuing the master's degree with the College of Computer Science and Technology, Shanghai University of Electric Power. His research interests include small target detection and transmission line abnormal target detection.

CHAO SUN was born in Jiangsu, China, in 1998. He received the master's degree from the College of Computer Science and Technology, Shanghai University of Electric Power, in 2022. His research interests include small target detection, traffic sign detection, and transmission line abnormal target detection.

MEIJING SHAN received the Ph.D. degree from East China Normal University, China, in 2009. She is currently an Assistant Professor with the East China University of Political Science and Law. She has published the results of her research in more than 20 papers in international journals, conference proceedings, and book chapters. Her research interests include cybercrime and computer forensics.