



## Article

# Transmission Lines Small-Target Detection Algorithm Research Based on YOLOv5

Qiuyan Cheng<sup>1</sup>, Guowu Yuan<sup>1</sup> , Dong Chen<sup>1</sup>, Bangwu Xu<sup>1</sup>, Enbang Chen<sup>2</sup> and Hao Zhou<sup>1,\*</sup> 

<sup>1</sup> School of Information, Yunnan University, Kunming 650504, China; chengqiuyan@itc.ynu.edu.cn (Q.C.); gwyuan@ynu.edu.cn (G.Y.); 12021215257@mail.ynu.edu.cn (D.C.); cheng18212703363@163.com (B.X.)  
<sup>2</sup> Kunming Enersun Technology Co., Ltd., Kunming 650500, China; ec249523687@163.com  
\* Correspondence: zhouhao@ynu.edu.cn

**Abstract:** The images captured using UAVs during inspection often contain a great deal of small targets related to transmission lines. These vulnerable elements are critical for ensuring the safe operation of these lines. However, due to various factors such as the small size of the targets, low resolution, complex background, and potential target aggregation, achieving accurate and real-time detection becomes challenging. To address these issues, this paper proposes a detection algorithm called P2-ECA-EIOU-YOLOv5 (P2E-YOLOv5). Firstly, to tackle the challenges posed by the issues of complex background and environmental interference impacting small targets, an ECA attention module is integrated into the network. The module effectively enhances the network's focus on small targets, while concurrently mitigating the influence of environmental interference. Secondly, considering the characteristics of small target size and low resolution, a new high-resolution detection head is introduced, making the network more sensitive to small targets. Lastly, the network utilizes the EIOU\_Loss as the regression loss function to improve the positioning accuracy of small targets, especially when they tend to aggregate. Experimental results demonstrate that the proposed P2E-YOLOv5 detection algorithm achieves an accuracy P (precision) of 96.0% and an average accuracy (mAP) of 97.0% for small-target detection in transmission lines.

**Keywords:** power line inspection; object detection; small targets; attention mechanisms; loss function



**Citation:** Cheng, Q.; Yuan, G.; Chen, D.; Xu, B.; Chen, E.; Zhou, H. Transmission Lines Small-Target Detection Algorithm Research Based on YOLOv5. *Appl. Sci.* **2023**, *13*, 9386. <https://doi.org/10.3390/app13169386>

Academic Editor: Nuno Silva

Received: 19 June 2023

Revised: 13 August 2023

Accepted: 14 August 2023

Published: 18 August 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The images captured using UAVs contain numerous small targets related to power transmission lines. These small objects, such as bolts, insulators, and various fittings, are often critical and vulnerable components that ensure the safe operation of power transmission and distribution lines. There are two ways of defining small targets in power line images. One approach is the relative size definition, where targets with an area less than 0.12% of the total image area are considered to be small targets (assuming a default input image size of  $640 \times 640$ ). The other approach is the absolute size definition. For instance, in COCO data, targets smaller than  $32 \times 32$  pixels are considered small targets [1]. Small targets are typically objects that are relatively small compared to the original image size. In these transmission line images, there are many small components, such as insulators capable of withstanding voltage and mechanical stress, as well as bolts used for fastening and connecting major components [2]. These components, found in the acquired power line images, exhibit characteristics such as small proportion, low resolution, easy aggregation, and indistinct features, categorizing them as small target objects. Additionally, power transmission lines are usually installed in complex natural environments, resulting in complex backgrounds in UVA-captured power transmission line images, which greatly interferes with the detection of small targets on the power lines. Therefore, improving the resolution of small transmission line targets in the images, enhancing their feature extraction, filtering complex backgrounds, and achieving high-precision identification and positioning for real-time and accurate detection pose challenging problems.

The object detection algorithm based on deep learning has shown great potential in enhancing object resolution, improving feature extraction, and achieving high-precision positioning in images. As a result, it has gained wide application in the domain of small-target detection in transmission lines. This algorithm can be broadly categorized into two main types:

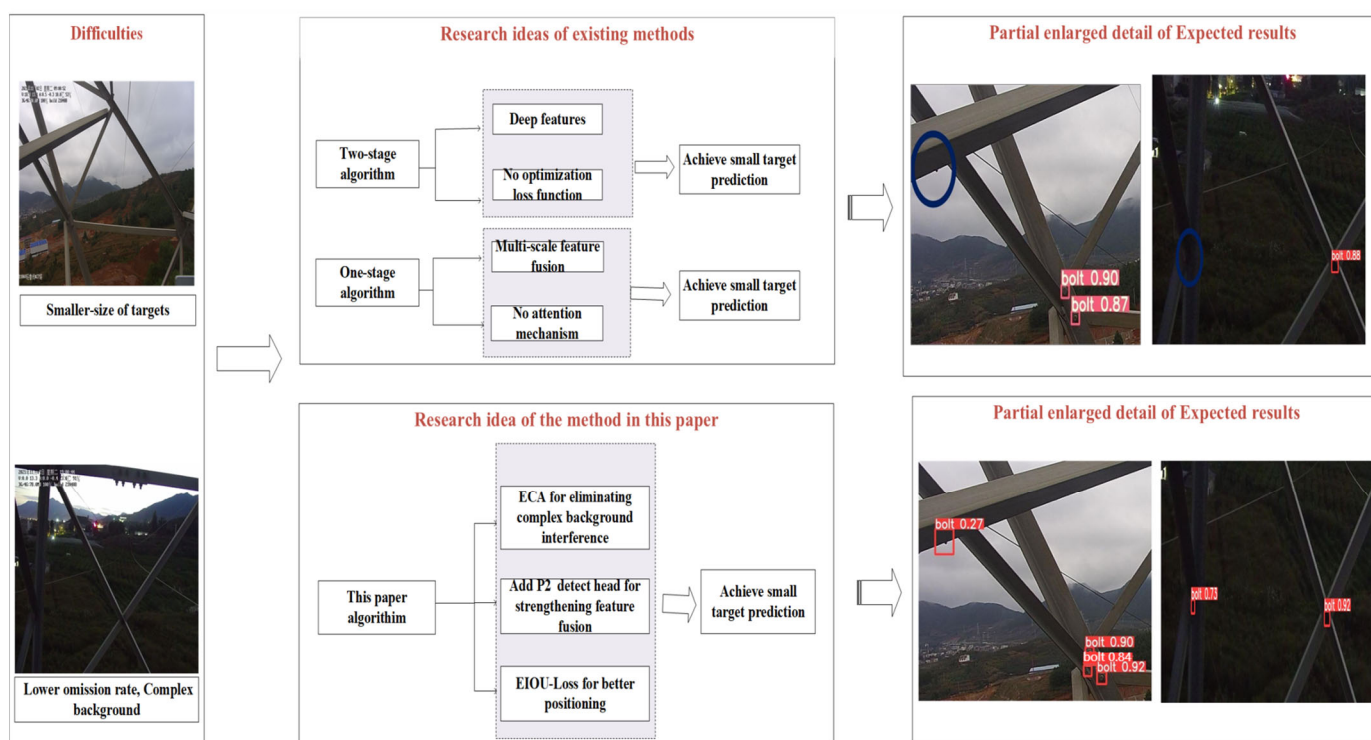
- (1) Two-stage detection algorithms, which are mainly represented by R-CNN [3], Fast-RCNN [4], Faster R-CNN [5], etc. Such algorithms usually generate target candidate regions and then extract features through convolution neural networks to achieve classification and localization predictions for target objects. In the literature [6], based on Faster R-CNN, a feature pyramid network (FPN) was proposed as a way of replacing the original RPN and fusing high-level features, rich in semantic information, with lower-level features, rich in positional information, in order to generate feature maps of different scales. Then, the target of the corresponding scale is predicted based on the feature map of different scales.
- (2) Single-stage detection algorithms, which are mainly represented by SSD [7], YOLO [8–11] series, etc. Such algorithms can achieve target classification and positioning prediction by directly extracting features through convolution neural networks. Redmon et al. proposed a YOLO [8] detection algorithm that divides the image into  $s \times s$  grids and directly predicts category probability and regression position information based on the surrounding box, which corresponds to each grid. This method does not generate candidate regions and improves the prediction speed. In the same year, Liu et al. [7] put forward the SSD algorithm, which draws on the idea of the YOLO algorithm and uses multi-scale learning to detect smaller targets on shallow-feature maps and larger targets on deeper-feature maps. Then, Ultralytics, a particle physics and artificial intelligence startup, proposed a single-stage object detection algorithm, YOLOv5 [11]. It uses a deep residual network to extract target features and combines the feature pyramid network FPN and perceptual adversarial network (PAN) [12] to efficiently fuse rich low-level and high-level feature information. It realizes multi-scale learning and effectively improves the detection performance of small targets.

In conclusion, the two-stage detection algorithm is capable of identifying and categorizing small targets using deep features. However, it fails to adequately capture the rich detailed characteristics of small targets present in shallow features. Moreover, deep features are obtained through multiple down-sampling, which often results in the loss of the fine details and spatial features of the small target, making it challenging to accurately determine their locations. On the other hand, the single-stage detection algorithm employs multi-scale feature fusion, which addresses the multi-scale issue. Nevertheless, in multi-scale feature fusion, the fusion of shallow features and deep features is not sufficient, which leads to the unsatisfactory detection accuracy of small targets. Additionally, small targets are often subject to complex backgrounds and susceptible to noise interference, limiting the algorithm's ability to eliminate such complex background interference. Taking into account the shortcomings and deficiencies of the existing methods, an improved method called P2E-YOLOv5 (P2-ECA-EIOU-YOLOv5) is proposed in this paper. This method not only fulfills the requirements of real-time detection but also achieves the highly precise detection of small targets in transmission lines. The specific implementation of this method is as follows:

1. To address the challenge of complex backgrounds and susceptibility to environmental interference in small targets, we enhance the network of capability to focus on these targets and mitigate environmental disturbances by incorporating the ECA (Efficient Channel Attention) [13] module. This addition improves the network's attention toward small targets, while simultaneously reducing the impact of environmental interference;
2. To cater to the imaging characteristics of small size and lower-resolution targets, a high-resolution P2 detection head is integrated into the network to enhance the detection ability for small targets;

3. Considering that small target features are not always obvious and tend to aggregate, we use the EIOU\_Loss [14] as the regression loss function of the network to improve the accurate identification and positioning of small targets.

In order to compare the characteristics of existing target detection algorithms and the proposed method more concisely, we have drawn a detailed research motivation diagram, as shown in Figure 1. Among them, the existing method uses Faster-RCNN [5] as the representative network framework, and the method adopted in this paper uses the improved YOLOv5 (P2E-YOLOv5) as the network framework. By randomly selecting four small target images of transmission lines for testing, the expected detection results are shown in Figure 1.



**Figure 1.** In the Figure, the presence of blue circles denotes the presence of diminutive undetected targets that eluded the grasp of the detection algorithm. Upon juxtaposing the anticipated outcomes of these two methodologies, it becomes conspicuously apparent that the suggested approach exhibits superior performance.

Furthermore, the proposed P2E-YOLOv5 detection algorithm in this study significantly enhances the resolution of small target feature maps of transmission lines. It effectively extracts relevant features of small targets and reduces the influence of environmental factors, making it highly suitable for small-target detection in transmission lines. Moreover, this algorithm demonstrates excellent detection performance when applied to other domains, such as facial defect detection and aerial images.

## 2. Related Work

Transmission line small-target detection is a challenging problem in the field of target detection, and it has gradually become a research hotspot both domestically and internationally in recent years.

Wu et al. [15] proposed a method for detecting small-target defects of transmission lines based on the Cascade R-CNN algorithm. Specifically, they used the ResNet101 (Residual Network) [16] network to extract target features and employed a multi-hierarchy detector to identify and classify small targets of transmission lines. Experimental results

demonstrate that this algorithm significantly improves the detection accuracy of small-target defects in transmission lines. Huang et al. [17] and others proposed an improved SSD small-target detection algorithm, named PA-SSD. They integrated deconvolution fusion units into the PAN algorithm to fuse feature maps of different scales, enhancing multi-scale feature fusion. Additionally, they established a new feature pyramid model by replacing feature maps in the original SSD algorithm. Experimental results show that the improved algorithm can extract feature maps with higher resolution, enhance feature extraction for small targets, and significantly improve the detection accuracy of power components in transmission lines while maintaining satisfactory detection speed.

Zou et al. [18] proposed an effective approach to expand the foreign body data set of transmission lines using scene enhancement, mix-up, and noise simulation. This method reduces the problem of limited pictures in the data set and enhances the diversity of the data set, leading to improved detection performance of the network on small targets. Li et al. [19] proposed a perceptive GAN method specifically designed for small-target detection. Their approach uses generators and discriminators to learn high-resolution feature representations of small targets from each other. Liu et al. [20] enhanced the feature extraction capability of the SSD network by replacing the original VGG network with the more powerful ResNet network. Furthermore, they integrated FPN into the network structure to achieve information fusion between upper and lower feature layers. The experimental results demonstrate significantly improved accuracy in detecting small and medium-scale targets in transmission line images compared to the original SSD algorithm. Li et al. [21] proposed an object detection algorithm based on the improved CenterNet. They built a multi-channel feature enhancement structure and introduced underlying details to improve the low detection accuracy caused by using a single feature in CenterNet. The research results show that the proposed algorithm achieves good detection results for power components and abnormal targets of transmission lines.

Additionally, Han et al. [22] proposed an insulator detection and defect identification algorithm based on YOLOv4. The enhanced algorithm utilizes feature multiplexing to improve the residual edge of the residual structure, significantly reducing the number of parameters and computational complexity of the model. The SA-Net (Random Attention Neural Network) [23] attention model is integrated into the feature fusion network to strengthen the focus on the target features and enhance their significance. Moreover, multiple outputs are added to the output layer to improve the model's capability to identify small, damaged insulator targets. Experimental results demonstrate that the enhanced algorithm significantly reduces the number of parameters while greatly improving the accuracy of insulators and defects detection.

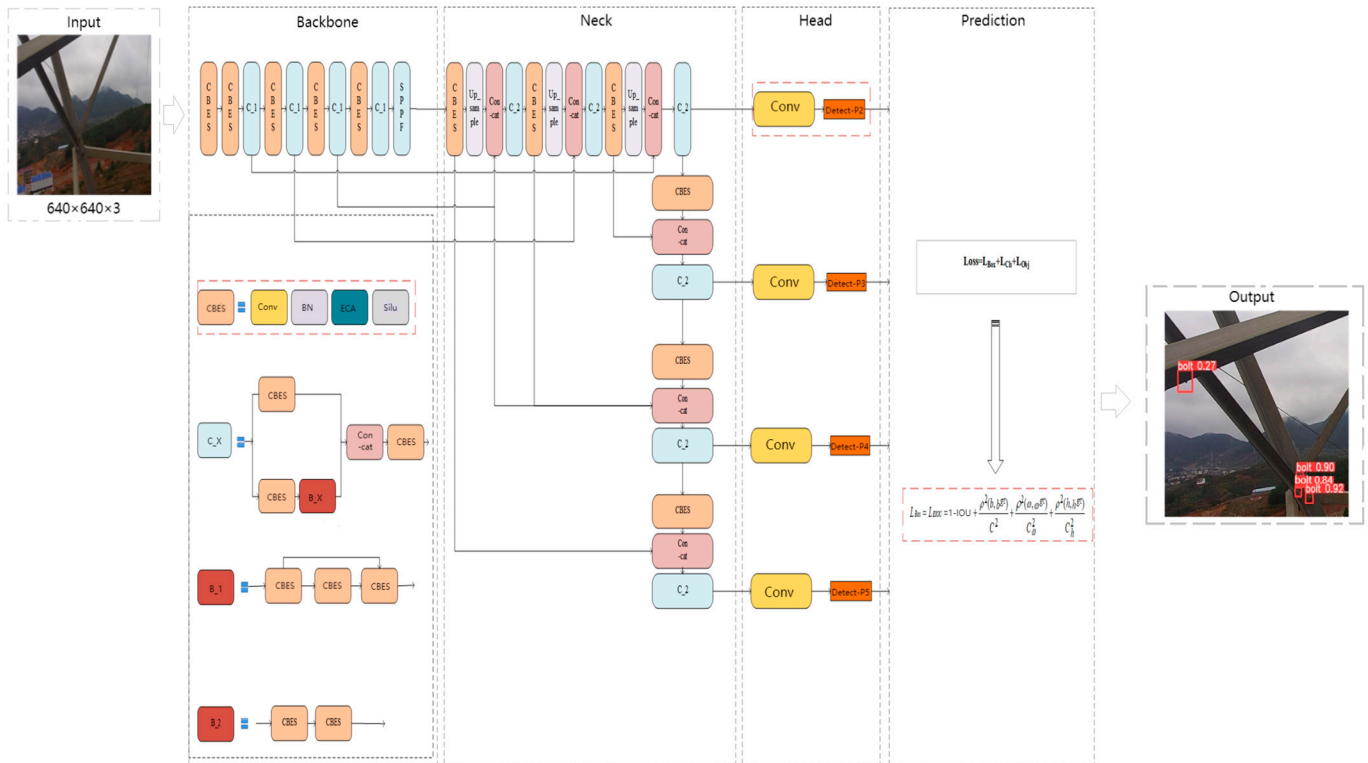
Recently, Su et al. [24] aimed to achieve high-precision identification and detection of small tower targets in remote sensing images. They conducted integrated modeling based on YOLOv5s and YOLOv5x algorithms and introduced the weighted boxes fusion (WBF) reasoning mechanism for model training and testing using a high-resolution remote sensing tower image data set. Additionally, they applied test-time augmentation (TTA) to the data set. Experimental results indicate that the integrated YOLOv5 model outperforms single model recognition in terms of accuracy. The model demonstrates excellent recognition capabilities and robustness even under complex backgrounds.

At present, Cao et al. [25] have proposed an automatic detection method for infrared small targets under complex backgrounds and conducted extensive experiments on the public data sets NUDT-SIRST and NUAA-SIRST. The results show that the proposed detection method exhibits excellent performance.

In summary, it is evident that the aforementioned research fails to achieve a balance between detection accuracy and detection speed, which is essential for small-target detection in transformation lines. The YOLOv5 detection model is widely used in small object detection tasks due to its high detection accuracy and fast detection speed. Therefore, this article aims to improve the YOLOv5 detection model and proposes the P2E-YOLOv5 detection method, which achieves high detection accuracy while maintaining real-time detection.

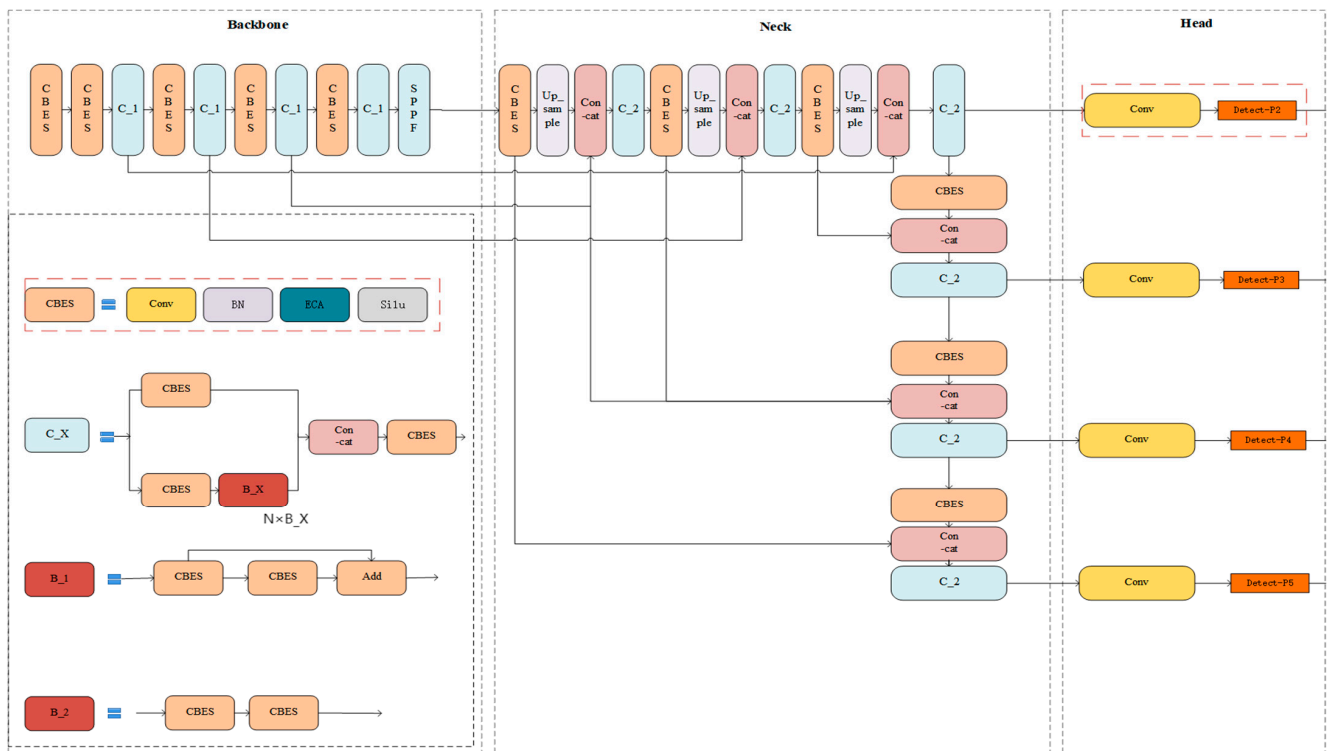
### 3. Introduction to the P2E-YOLOv5 Network

To achieve accurate and efficient detection of small targets in transmission lines, this paper proposes the adoption of the YOLOv5s network model with a simplified architecture. The P2E-YOLOv5 network, as depicted in Figure 2, represents the overall fundamental structure of our proposed approach. The main network structure of the P2E-YOLOv5 network is shown in Figure 3.



**Figure 2.** The architectural framework of P2E-YOLOv5 encompasses several crucial components, namely Input, Backbone (comprising the backbone network), Neck (housing the neck network), Head (incorporating the detection head), Prediction, and Output. Notably, this paper introduces certain modifications denoted by a red rectangular dotted line box.

To address the challenges posed by the complex background and vulnerability to environmental factors of small targets in transmission lines, this paper introduces the use of an ECA (Efficient Channel Attention) mechanism. The CBES module is constructed by incorporating an ECA attention module after the BN (Batch Normalization) layer of the CBS module, which is widely distributed in the backbone and neck parts (see Figure 5). This integration encourages the network to focus more on the small target area, reducing interference from complex backgrounds and enhancing the network’s ability to extract features relevant to small targets. Considering the characteristics of small targets in transmission lines, such as their small size and low resolution, a P2 detection head is added to the Head part. This head has a high resolution of  $160 \times 160$  pixels and facilitates the effective fusion of rich low-level features and deep semantic information. It is generated by low-level and high-resolution feature maps, making it highly sensitive to the detection of small targets. To address the challenge of small targets being indistinct and prone to aggregation, the EIOU\_Loss is considered as the loss function of network regression in the prediction part. This choice effectively enhances the regression accuracy of the bounding box, thereby improving the precise positioning and identification of small targets by the network.



**Figure 3.** The main network structure of P2E-YOLOv5 network consists of the following three parts: Backbone (backbone network), Neck (neck network), and Head (detection head).

### 3.1. High-Resolution Small-Target Detection Head

To address the challenges posed by the small target size and low resolution of the transmission line, this paper proposes the addition of a high-resolution P2 detection head to the Head part of YOLOv5. This enhancement aims to strengthen the feature fusion between shallow features of small targets and deep semantic information, thereby improving the feature extraction ability for small targets.

In the original YOLOv5 network model, the backbone network obtains 5-layer (P1, P2, P3, P4, P5) feature expressions after 5 down-sampling operations, with  $P_j$  indicating a resolution of  $1/2^j$  of the input image. In the Neck part of the network, top-down FPN and bottom-up PAN structures are used to achieve feature fusion at different scales, and object detection is performed using the detection head on the three-level feature map P3, P4, and P5.

However, when dealing with the small-target data set in transmission lines, the pixel scale of these targets is often less than  $32 \times 32$ . As a result, after multiple down-sampling operations, most of the detailed features and spatial information are lost, making it challenging for the higher resolution P3 layer detection head to effectively detect these small targets.

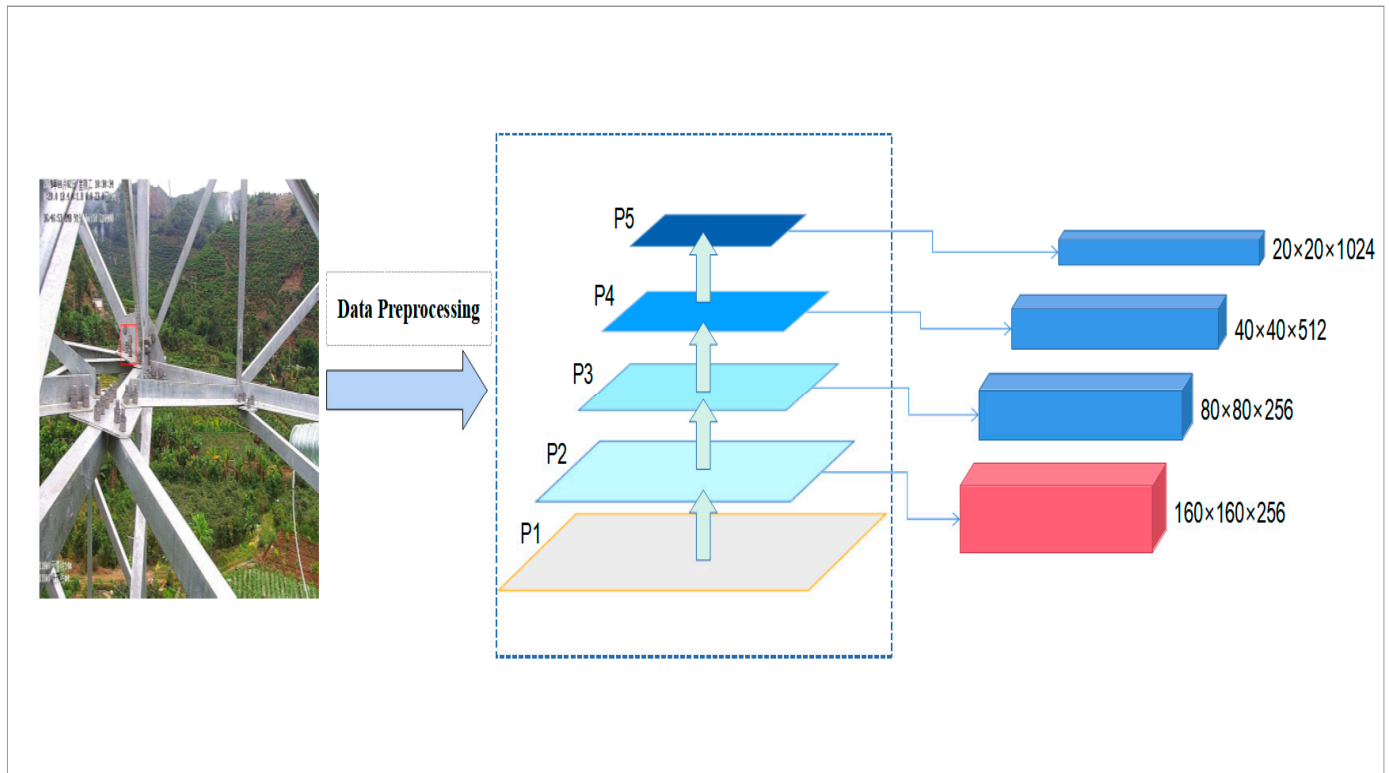
To achieve the detection of the aforementioned small targets, a new detection head is introduced in the P2 layer feature map of the YOLOv5 network model, as illustrated in Figure 4.

Firstly, the detection head has a high resolution of  $160 \times 160$  pixels, enabling it to detect tiny targets as small as  $8 \times 8$  pixels. This high resolution makes it highly sensitive to smaller targets, enhancing the ability to detect them accurately. The new high-resolution P2 detection head only requires two down-sampling operations on the backbone network, retaining extremely rich shallow feature information of small targets.

Secondly, in the Neck part, the top-down features of the P2 layer are merged with the same-scale features of the backbone network, allowing for the fusion of shallow features

and deep semantic features. This fusion process results in output features that effectively combine multiple input features.

Lastly, the integration of the P2 layer detection head with the original three detection heads effectively addresses the issue of scale variance. The P2 layer detection head is derived from low-level, high-resolution feature maps, which contain abundant detail features and semantic information related to small targets.



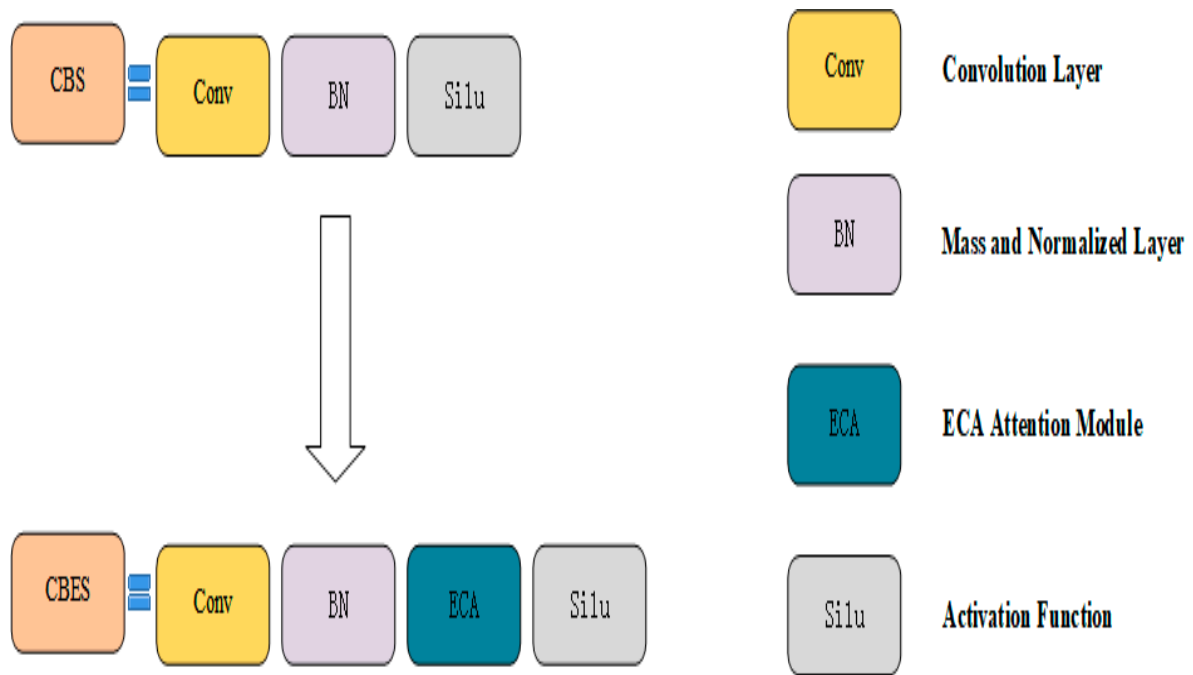
**Figure 4.** By leveraging the three detection heads derived from the initial P3, P4, and P5 layer feature maps, a novel detection head with an elevated resolution of  $160 \times 160$  pixels is extracted from the P2 layer feature map.

As a result, this detection head becomes highly sensitive to detecting small targets in transmission lines. Although incorporating this detection head increases the computational complexity and memory requirements of the model, the substantial enhancement it brings to the detection accuracy of small targets justifies its inclusion.

### 3.2. ECA Attention Mechanism

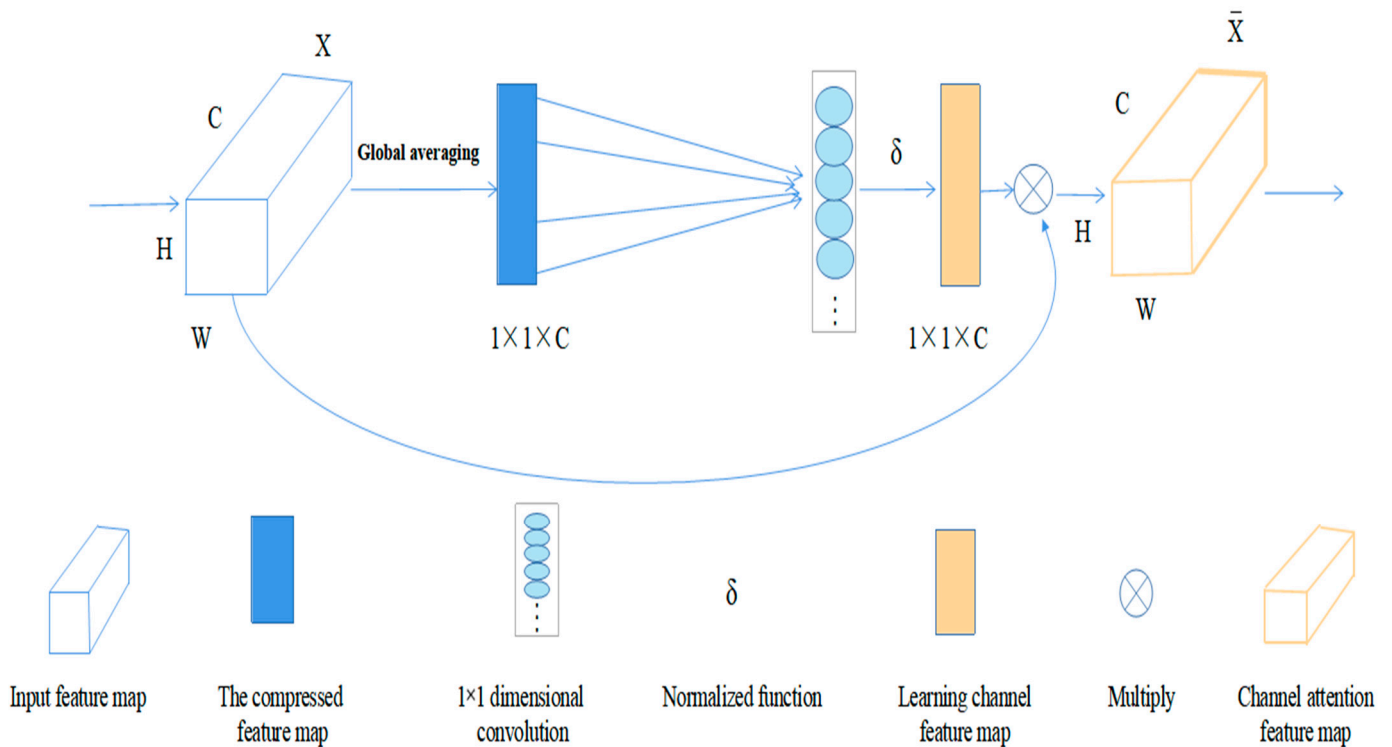
In the case of small target images of transmission lines with complex environmental backgrounds, such as insulators and bolts, there are challenges like inconspicuous features and high susceptibility to environmental interference. These factors make it difficult for the detection model to accurately identify them. To address these issues, this paper introduces the ECA attention mechanism, as depicted in Figure 5, to effectively mitigate the impact of environmental factors and improve the network's ability to focus on small target areas, thus enhancing the accuracy of detection.

By incorporating the ECA attention module after the BN layer of the CBS module, the interference of environmental factors on small targets can be significantly reduced. This allows the model to effectively identify obstructed small targets and strengthen its focus on the small target area. Furthermore, the CBES module composed of the ECA attention module can be added to both the backbone network and neck network. This enhancement strengthens the network model's feature extraction capability for small targets, ultimately leading to improved detection performance.



**Figure 5.** The ECA attention mechanism is added after the BN layer of the CBS module.

The ECA attention module is a lightweight module that can be seamlessly integrated into the CNN architecture. It employs a local cross-channel interaction strategy without dimension reduction, effectively preventing any negative impact on channel attention learning caused by dimension reduction. This proper cross-channel interaction allows for a significant reduction in model complexity while preserving performance. The structure of the ECA module is illustrated in Figure 6:



**Figure 6.** Specific composition of ECA attention module.



The ECA attention module is described in detail as follows:

- (1) The first input is an  $H \times W \times C$  dimensional feature graph;
- (2) The input feature map undergoes spatial feature compression: in the spatial dimension, a feature map of  $1 \times 1 \times C$  is obtained by using global average pooling;
- (3) Channel features are learned for the compressed feature graph: the importance of different channels is learned through  $1 \times 1$  convolution, and the output dimension remains  $1 \times 1 \times C$ . Among them,

$$k = \psi(C) = \left\lfloor \frac{\log 2C}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}}. \quad (1)$$

Formula: the size of  $k$  can be obtained adaptive, where  $C$  represents the channel dimension,  $\lfloor t \rfloor_{\text{odd}}$  represents the odd number closest to  $t$ , where  $\gamma = 2, b = 1$ .

- (4) Finally, the channel attention is combined, and the feature map of channel attention  $1 \times 1 \times C$  is multiplied channel by channel with the original input feature map  $H \times W \times C$  to generate a feature map with channel attention.

### 3.3. Optimize the Loss Function

In order to address the challenges posed by the indistinct and aggregated characteristics of small targets on transmission lines, this paper proposes the use of EIOU\_Loss as the loss function for network regression. This approach aims to improve the regression accuracy of bounding boxes, enhancing the network's ability to accurately identify and localize small targets.

Small-target images of transmission lines captured using UAV, such as insulators and bolts, constitute a very small proportion of the entire image, and their low resolution result in limited visual information, making it challenging to extract clear and distinct features. These characteristics pose difficulties for detection models in accurately locating and identifying these small targets. Additionally, when small targets are densely clustered, they may merge into a single point in the deep feature map after multiple down-sampling, causing the detection model to fail to distinguish them. Furthermore, when small targets are too close to each other, the non-maximum suppression operation in post-processing may filter out a significant number of correct predictional bounding boxes, resulting in missed detections.

The original YOLOv5 network used CIOU\_Loss [24] as the loss function for network regression. Although CIOU\_Loss considered the overlap area, center point distance, and aspect ratio of bounding box regression, it only accounted for the aspect ratio difference in bounding boxes and did not consider the actual differences between width and height and their confidence. Additionally, when the aspect ratio of the predicted box changes proportionally, the aspect ratio penalty in CIOU\_Loss loses its effectiveness, making it challenging to accurately locate and recognize small targets. To address these issues, this paper adopts the EIOU\_Loss function as the regression loss function for the P2E-YOLOv5 model. EIOU\_Loss resolves the ambiguity in the definition of aspect ratio and the proportional change of aspect ratio that affects the penalty in CIOU\_Loss. By doing so, the EIOU\_Loss effectively improves the detection model's precise positioning of small targets such as insulators and bolts. This enhancement leads to improved convergence and significantly reduces the model's omission rate for these small targets.

The EIOU\_Loss function used in this paper consists of three components: overlap loss  $L_{\text{IOU}}$ , center distance loss  $L_{\text{dis}}$ , and width and height loss  $L_{\text{asp}}$ . The first two parts retain the benefits of the CIOU method, while the width and height loss divides the aspect ratio loss term into the difference between the predicted width and height and the minimum width and height of the enclosing box. This modification is applied to the penalty term of CIOU to expedite model convergence and enhance regression accuracy. The calculation formula for EIOU\_Loss is as follows:

$$L_{EIOU} = L_{IOU} + L_{dis} + L_{asp} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{C^2} + \frac{\rho^2(\omega, \omega^{gt})}{C_\omega^2} + \frac{\rho^2(h, h^{gt})}{C_h^2} \quad (2)$$

where: IOU represents the intersection over union ratio between the prediction box and the ground truth box,  $b$  and  $b^{gt}$  represent the center points of the prediction box and ground truth box,  $\rho(b, b^{gt})$  represents the Euclidean distance between the center point of the prediction box and the ground truth box,  $C$  represents the diagonal distance of the smallest enclosing rectangle of prediction box and ground truth box, and  $\omega^{gt}$  and  $h^{gt}$  are respectively the width and height of the ground truth box.  $\Omega$  and  $h$  are the width and height of the prediction box, respectively, while  $C_\omega$  and  $C_h$  are the width and height of the minimum bounding box that includes the ground truth box and prediction box.

#### 4. Experimental Results and Analysis

The hardware and software platform configurations used in this experiment are as follows: Intel(R)Core (TM)i7-6700 CPU @ 3.40Ghz, NVIDIA TITAN X(Pascal); Operating system: Windows10 64-bit operating system; CUDA version is 11.0, the python version is 3.7 and the deep learning framework is Pytorch 1.7.1.

##### 4.1. Evaluation Indicators and Data Set Preparation

The following metrics were utilized in this study: Accuracy (P); Recall (R); Average Precision (AP); Mean Average Precision (mAP), and frames per second (FPS). These metrics were employed to evaluate the model from different perspectives. The calculation expressions of P, R, AP, and mAP are presented in Table 1.

**Table 1.** In this Table,  $N_{TP}$ ,  $N_{FP}$ , and  $N_{FN}$  represent the number of samples predicted correctly, the number of samples predicted incorrectly, and the number of samples not recognized, and  $N$  denotes the total number of categories specified by the model.

P	R	AP	mAP
$P = \frac{N_{TP}}{N_{TP} + N_{FP}}$	$R = \frac{N_{TP}}{N_{TP} + N_{FN}}$	$AP = \int_0^1 P(R) dR$	$mAP = (\sum_i^N AP_i) / N$

The original data used in this experiment is sourced from a power grid company in Yunnan, which provides a small-target data set of transmission lines. This data set contains various small targets of transmission lines, such as insulators and bolts. To augment the data, random horizontal flips, rotation transformations, brightness adjustments, and other data expansion techniques were applied to the original set of 1076 images, resulting in a new data set containing 3200 images. The new data set was annotated using image labeling software labeling and split into the training, validation, and test sets with an 8:1:1 ratio. For training, 300 epochs were used with an initial learning rate of 0.01 and a weight decay coefficient of 0.0005. The same data set, namely the small-target data set of transmission lines, was utilized for both training and testing in this experiment.

##### 4.2. Experiment

In this experiment, the small-target data set of transmission lines was subjected to data processing, and the analysis is presented in Figure 7.

In Figure 7, the darker the blue area, the greater the concentration of small targets. The left figure displays the distribution of the center coordinates for small target objects in the transmission line’s small-target data set. The coordinates are relative coordinates, which are relative to the coordinates of the whole transmission line small target image, and its size range is mapped to (0, 1). The central coordinates of the small target are observed to be mainly concentrated in the range (0.4~0.6, 0.0~1.0). The right figure displays the height and width distribution of small target objects relative to the whole image. It is evident that the small target objects are significantly smaller compared to the entire image. In fact, a

majority of small target objects account for less than 0.12% of the whole image area (the definition of relative size for small targets is elaborated in the introduction Section 1). This indicates that most of the targets in this data set are indeed small targets.

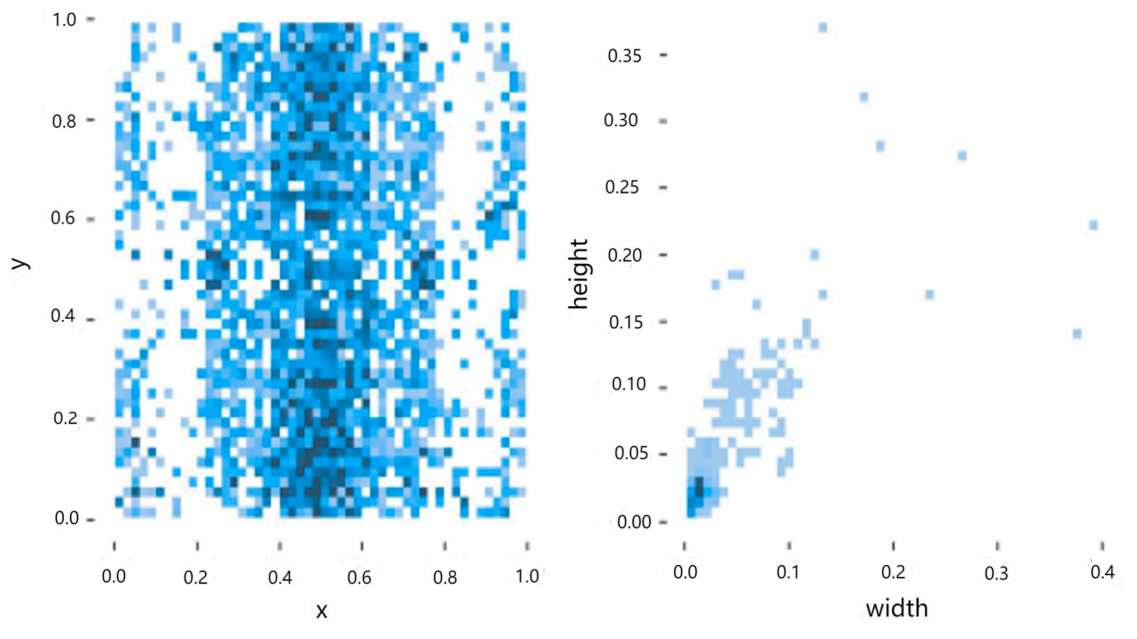


Figure 7. Among them, the horizontal direction is represented by the X-axis and the vertical direction is represented by the Y-axis. Data analysis of transmission line small-target data set.

Subsequently, training experiments were conducted on the small-target data set of transmission lines, and the original YOLOv5 training results were compared with the improved P2E-YOLOv5 training results:

The test results before the algorithm improvement are shown in Figure 8:

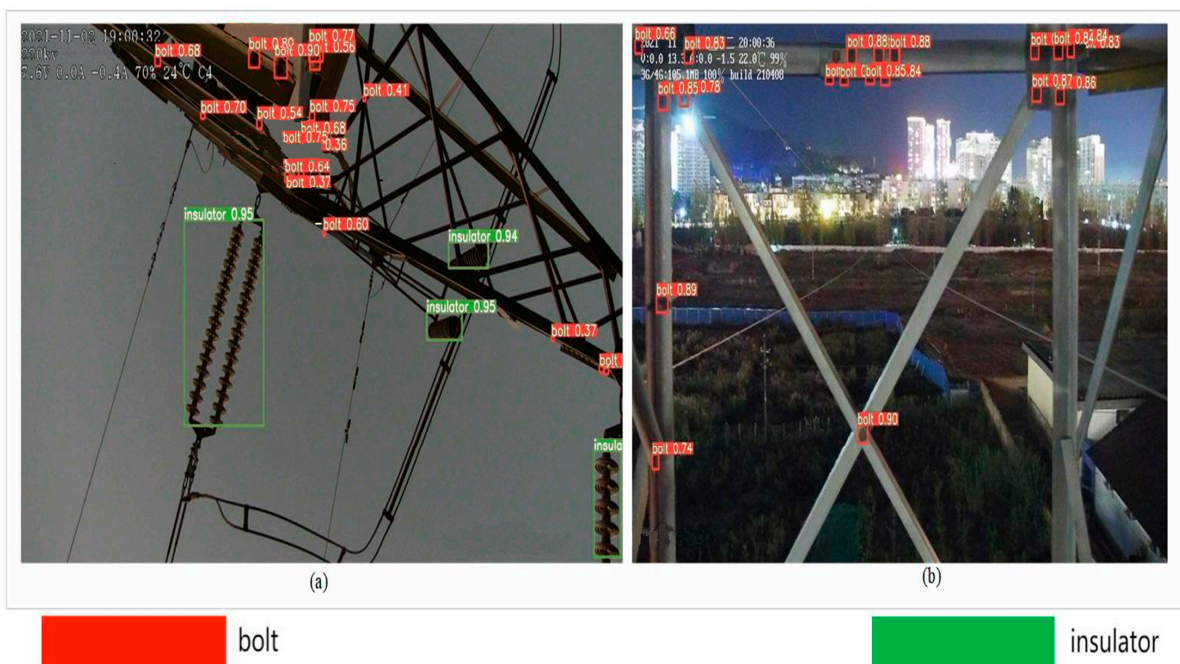


Figure 8. The test results prior to enhancement depict (a,b) as a pair of randomly selected images for the purpose of evaluation.

The improved model test results are shown in Figure 9:

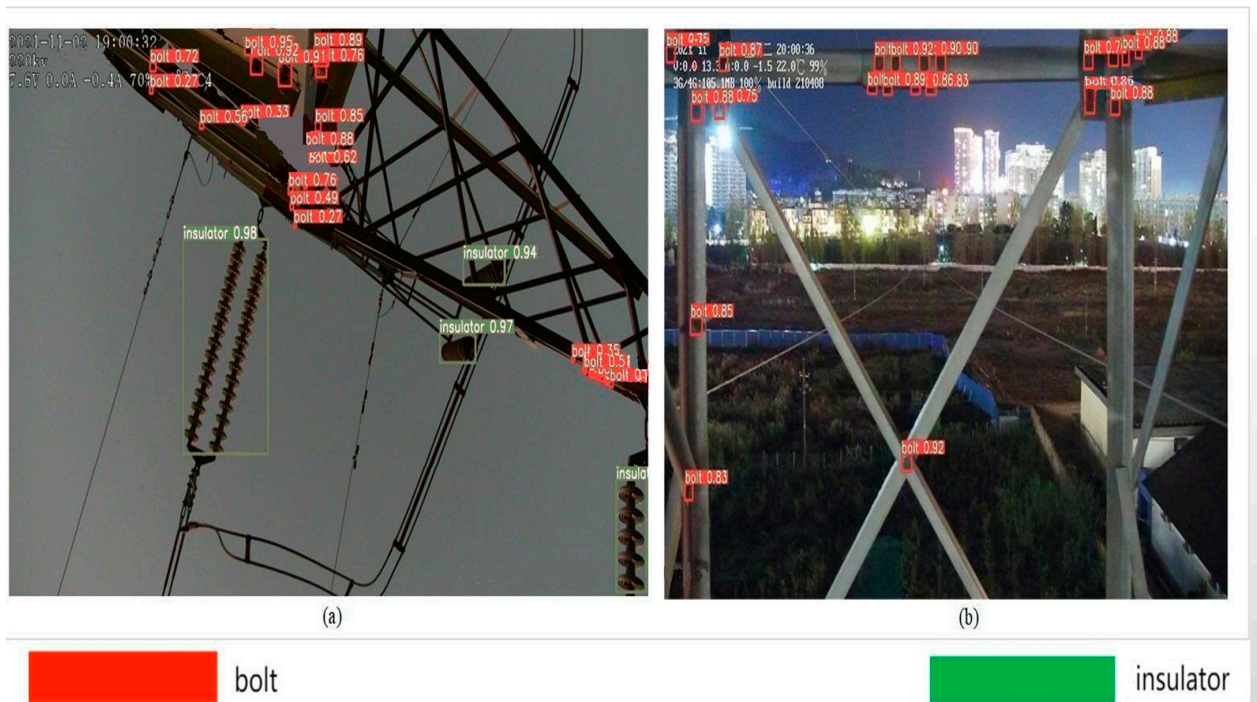


Figure 9. Improved test results; (a,b) represent two random test images after model enhancement.

Model performance comparison results are shown in Figures 10–13:

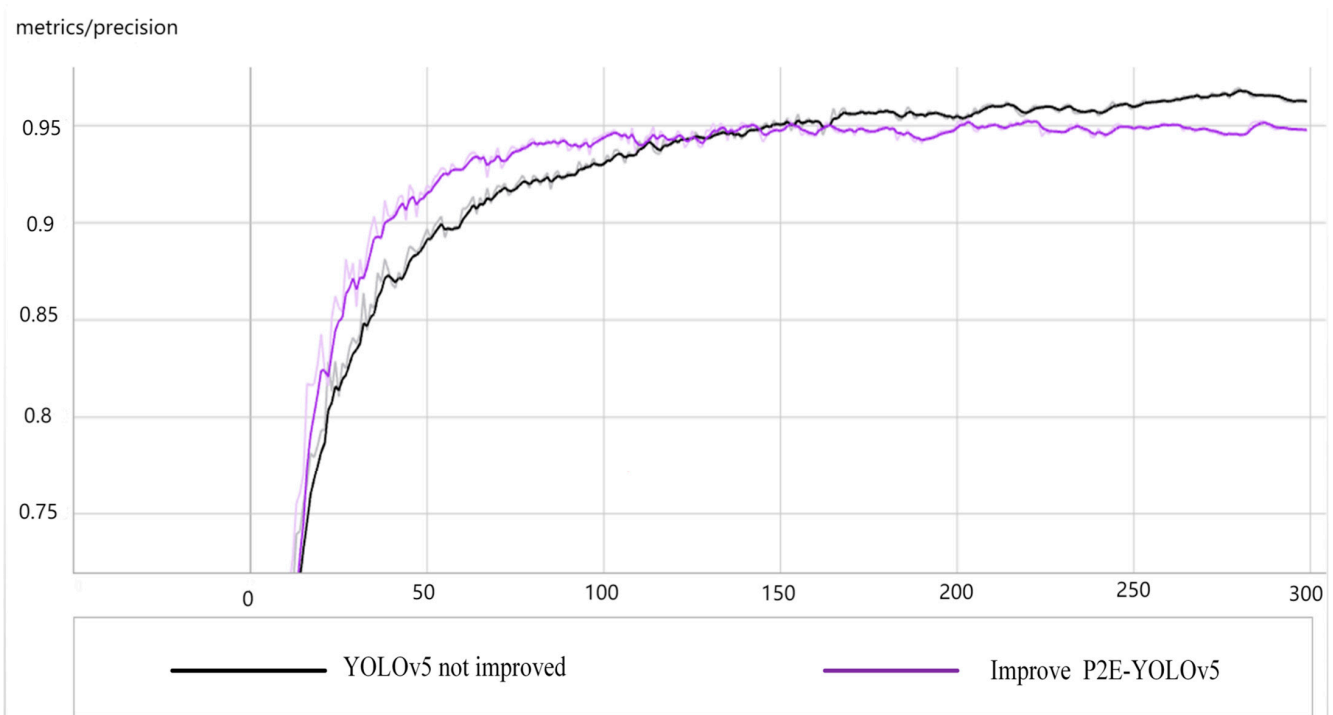
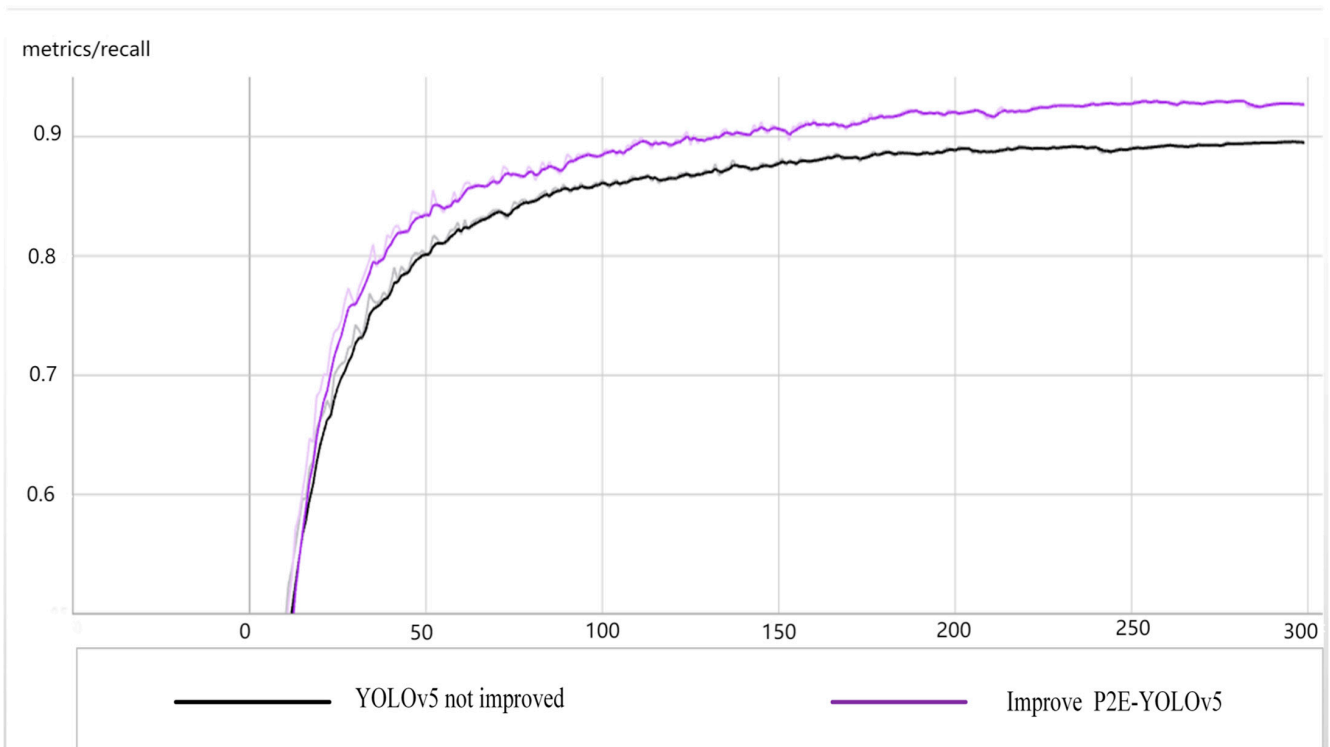
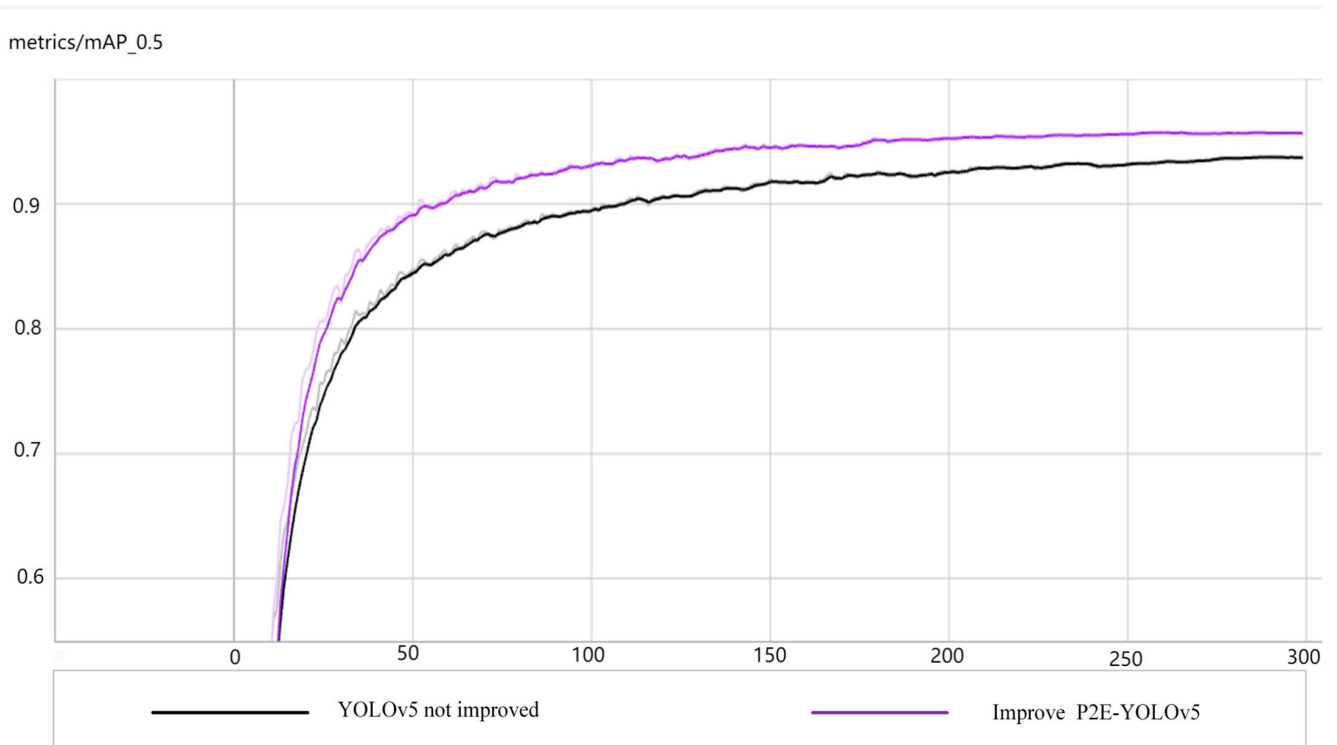


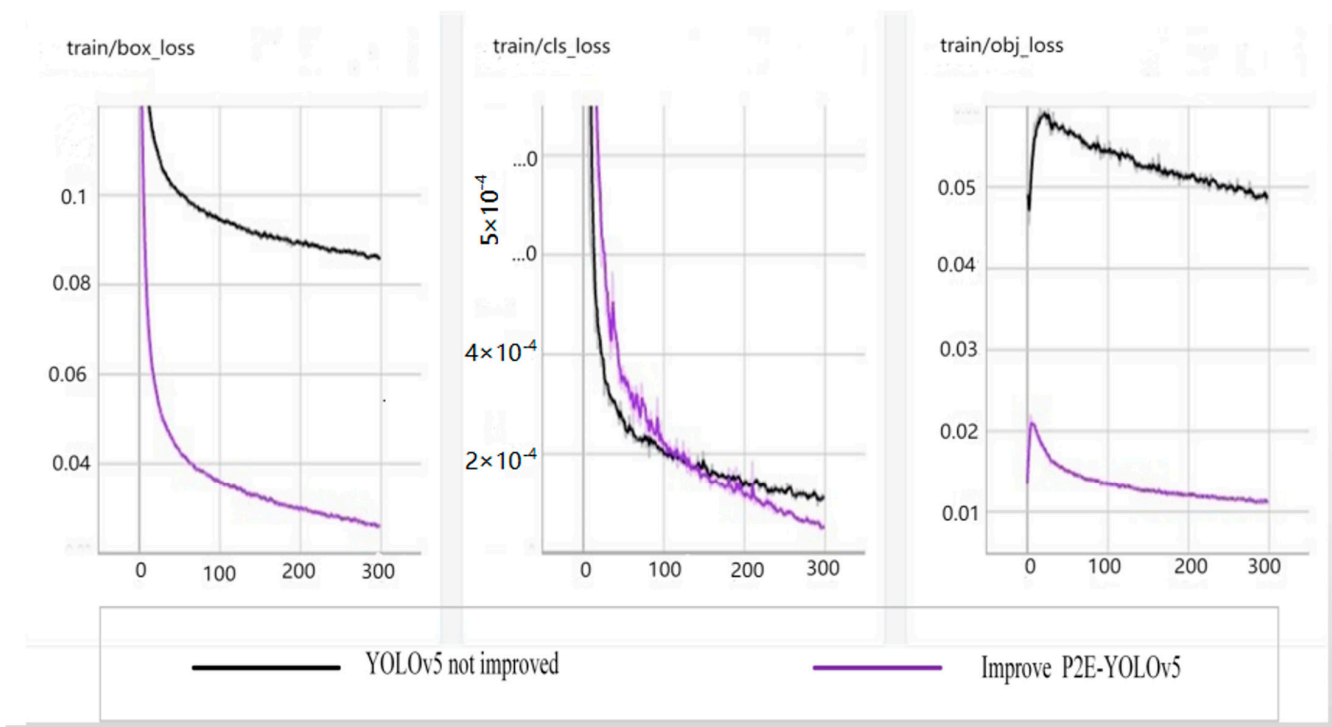
Figure 10. Among them, the horizontal direction is the X-axis and the vertical direction is the Y-axis. Comparison results of accuracy.



**Figure 11.** Among them, the horizontal direction is the X-axis and the vertical direction is the Y-axis. Comparison results of recall rates.



**Figure 12.** Among them, the horizontal direction is the X-axis and the vertical direction is the Y-axis. Average precision mAP comparison results.



**Figure 13.** Among them, the horizontal direction is represented by the X-axis and the vertical direction is represented by the Y-axis. Comparative results of training losses. It is evident that the regression loss, classification loss, and confidence loss of the improved YOLOv5 network converge faster during model training.

From the above comparison results, it is evident that the improved P2E-YOLOv5 outperforms the original YOLOv5 in terms of accuracy, recall rate, average accuracy, and loss index. The improved model shows a significant enhancement in detecting objects with low resolution and small sizes in images.

#### 4.3. Comparative Experiment

To demonstrate the effectiveness of the proposed model, we conducted a comparative experiment using the existing small-target data set of transmission lines. We selected Faster-RCNN, RetinaNet, SSD, YOLOv3\_spp, YOLOX, and YOLOv5 as the comparison network model. The experimental results are presented in Table 2.

**Table 2.** Comparison of experimental results of different network models. Additionally, mAP\_0.5 represents the mean precision at an intersection ratio of 0.5.

Model	mAP_0.5 (%)	FPS (Frames)
Faster-RCNN	87.2	3.968
RetinaNet	78.3	12.531
SSD	53.5	76.335
YOLOv3_spp	79.8	49.261
YOLOX	90.4	25.031
YOLOv5	93.7	101.010
P2E-YOLOv5 (Ours)	97.0	87.719

Table 2 clearly shows that the Mean Average Precision (mAP), which reflects the model's detection performance, exhibits significant improvements with the optimized P2E-YOLOv5 when compared to various other methods. Specifically, compared to Fast RCNN, P2E-YOLOv5 shows a 9.8% improvement in mAP. The improvement over RetinaNet is 18.7%, and over SSD, it is 43.5%. In comparison to the YOLOv3\_spp, P2E-YOLOv5 achieves

a 17.2% higher mAP. Additionally, compared to YOLOX, P2E-YOLOv5 demonstrates a 6.6% mAP improvement after optimization. Even when compared to the original YOLOv5, P2E-YOLOv5 still shows an overall mAP improvement of 3.3%. These results indicate that the proposed P2E-YOLOv5 detection method in this paper outperforms other detection algorithms significantly in terms of detection accuracy.

Furthermore, in terms of model detection speed, the optimized P2E-YOLOv5 achieved an impressive 87.719 frames per second, second only to the detection speed of the original YOLOv5. The slight decrease in detection speed compared to the original YOLOv5 is primarily due to the additional computation introduced by the newly added P2 detection head, which slightly increases the model's reasoning time. Nevertheless, when compared to other detection methods, P2E-YOLOv5 maintains a significant advantage, meeting the requirements of real-time detection.

Additionally, to demonstrate the effectiveness of the proposed method in this paper, we compared it with the latest infrared small object detection method [25] through comparative experiments on the same public data sets, NUDT-SIRST and NUAA-SIRST.

The experimental results demonstrate that compared to the infrared small object detection method, the proposed method improves the recall rate (R) by 1.2% and 2% on NUST-SIRST and NUAA-SIRST, respectively. In addition, the method achieves an increase of 42.493 frames per second (FPS) on NUAA-SIRST, indicating a faster detection speed.

In conclusion, the P2E-YOLOv5 algorithm proposed in this paper outperforms other target detection algorithms in all aspects. It effectively addresses the challenge of achieving high detection accuracy for small-target detection in transmission lines and fulfills the need for real-time detection.

#### 4.4. Ablation Experiment

The enhancement effect of each new or improved module on the overall model was explored through an ablation experiment using the transmission line small-target data set. Starting with the original YOLOv5, each module was successively added for experiments until the final model, P2E-YOLOv5, was obtained. The experimental results are presented in Table 3.

**Table 3.** Influences of different new modules on the model.

Model	P (%)	R (%)	mAP_0.5 (%)	FPS (Frames)
YOLOv5s	95.7	91.2	94.7	101.010
YOLOv5s + P2	95.5	94.7	96.9	86.207
YOLOv5s + P2 + ECA	95.9	95.0	96.9	88.495
YOLOv5s + P2 + ECA + EIOU	96.0	94.7	97.0	87.719

As shown in Table 3, the addition of a small-target detection head result in a slight decrease in the model's accuracy (P), but it significantly improved the recall rate (R) and mean average accuracy (mAP) by 3.5% and 2.2%, respectively. It indicates that the inclusion of the P2 layer small-target detection head greatly enhances the model's performance, effectively improving its ability to detect small targets. Subsequently, the ECA attention module was added on top of the small-target detection heads. While the mean average accuracy remained relatively unchanged, the accuracy (P) and recall rate (R) of the model increased by 0.4% and 0.3%, respectively. This demonstrates that the ECA attention module effectively enhances the network's focus on small target objects and improves the model's feature extraction capability for small targets. Finally, the EIOU\_Loss function was added on top of the small-target detection head and ECA attention module. Although the recall rate (R) of the model decreased, both the accuracy (P) and mean average accuracy (mAP) were improved. This indicates that the EIOU\_Loss effectively enhances the model's detection accuracy for small targets.

Furthermore, in terms of detection aging, the inference detection speed of the model decreased due to the increased computational load caused by the large detection layer scale.

However, after the addition of the ECA attention module and EIOU\_Loss function, there was a noticeable increase in the detection speed.

The data analysis results mentioned above align well with the theoretical analysis, providing further validation of the algorithm’s rationality and effectiveness proposed in this paper. The approach not only meets the requirements of real-time detection but also significantly improves the detection accuracy of small targets.

#### 4.5. Detection Effect

To validate the effectiveness of the P2E-YOLOv5 model for small-target detection in transmission lines, we utilized the trained model on the augmented small-target data set to detect images from the test set. Furthermore, we randomly selected three transmission line images captured in different scenes for detection. The test results are depicted in Figure 14.

In Figure 14, the test results of each algorithm are shown with rectangle boxes marking the successfully detected small targets. Additionally, blue circles indicate small missed targets that were missed by the detection algorithm.

Upon observation, the Faster-RCNN and YOLOX detection algorithms demonstrate high detection accuracy, accurately identifying most small target objects. However, for some small targets with subtle features, some cases are missed. RetinaNet, SSD, and YOLOv3\_spp can detect most small targets, but with low accuracy and numerous omissions. The YOLOv5 detection algorithm shows high detection accuracy and can detect most small targets, yet some cases are still missed.

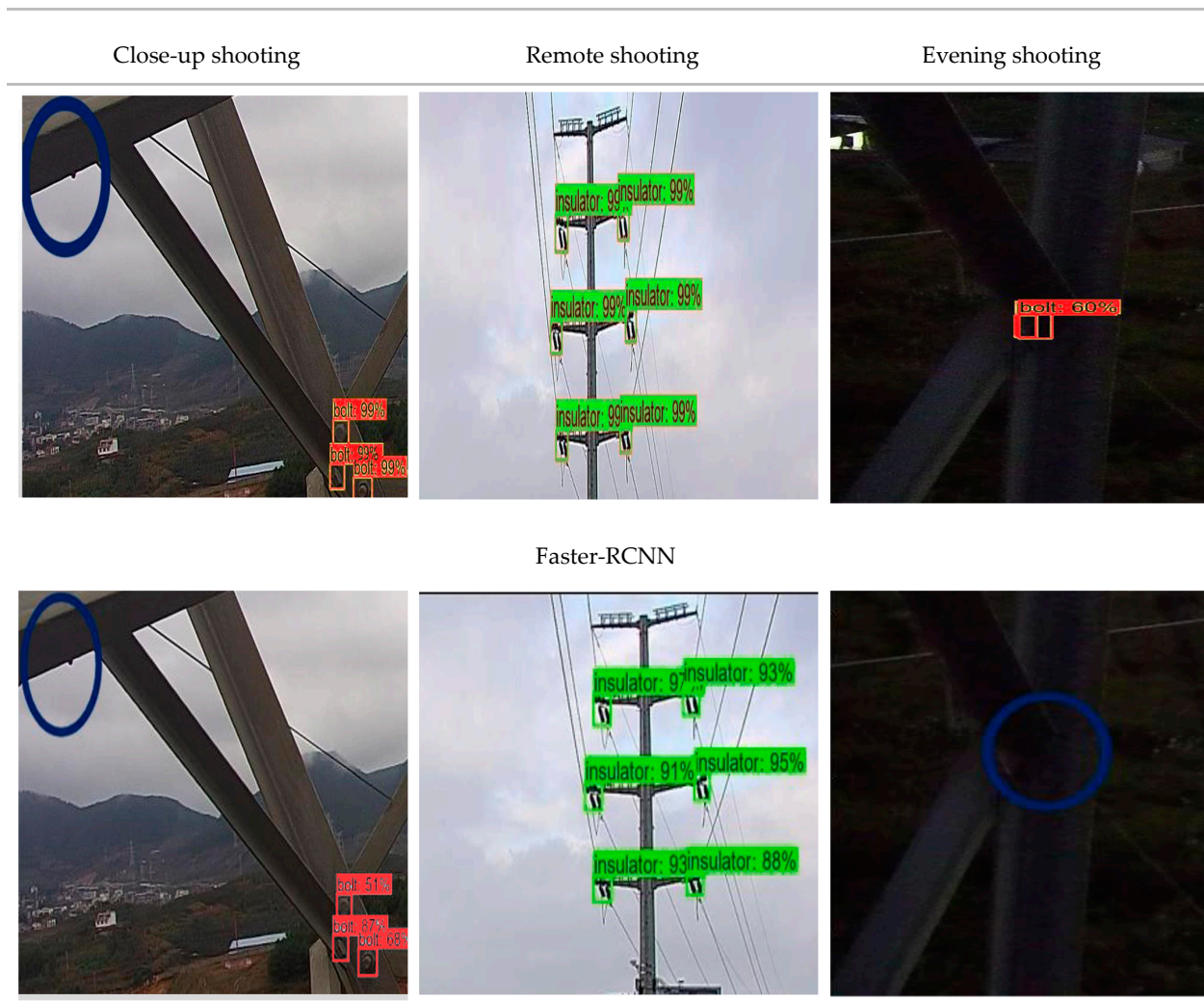


Figure 14. Cont.





Figure 14. Cont.



**Figure 14.** This figure presents a detailed comparison of the detection performance of various algorithms, including Faster-RCNN, RetinaNet, SSD, YOLOv3\_spp, YOLOX, YOLOv5, and P2E-YOLOv5, on small targets in transmission line images captured in different scenes.

Comparing the six detection algorithms, the P2E-YOLOv5 algorithm stands out for accurately identifying small targets in different scenes with higher detection accuracy and the lowest missing rate. Several factors contribute to this success. Firstly, the added ECA attention module significantly enhances the network’s focus on small targets and reduces the impact of environmental interference. Secondly, the newly introduced high-resolution P2 detection header significantly improves the network’s ability to detect small targets. Lastly, employing EIOU\_Loss is used as the regression loss function to enhance the precise identification and positioning of small targets. Consequently, the P2E-YOLOv5 detection algorithm outperforms the other six detection algorithms on a small-target detection data set of transmission lines.

### 5. Conclusions

In response to the challenging task of accurately and efficiently detecting small targets in transmission lines, this paper proposes a novel detection method called P2E-YOLOv5. This method incorporates the ECA attention mechanism into the network, introduces high-resolution detection heads that exhibit higher sensitivity to small targets, and optimizes

the regression loss function using EIOU\_Loss. These improvements significantly enhance the detection efficiency and capability of small targets amidst complex backgrounds. We evaluate the performance of our algorithm on a data set specially designed for small-target detection in power lines, and our findings demonstrate that the proposed method outperforms other detection algorithms, showcasing superior detection performance. Despite a slight increase in memory usage due to the addition of high-resolution detection heads, the method maintains a high level of detection accuracy. Considering the trade-off between detection accuracy and speed, our P2E-YOLOv5 detection method proves to be the optimal choice. For future work, our goal is to further reduce the computational costs associated with P2 detection heads while maintaining comparable detection accuracy. Additionally, we aim to continue improving the network's detection speed to enhance overall performance.

**Author Contributions:** Conceptualization, H.Z., G.Y. and Q.C.; methodology, Q.C. and E.C.; software, Q.C.; validation, B.X. and D.C.; investigation, Q.C., D.C. and B.X.; resources, D.C.; writing—original draft preparation, Q.C.; writing—review and editing, H.Z. and Q.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Major Science and Technology Project in Yunnan Province, grant number 202202AD080004. The APC was funded by Yunnan Province Science and Technology Department.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Restrictions apply to the availability of these data. Data were obtained from Yunnan Power Grid Co., Ltd. and are available from the authors with the permission of Yunnan Power Grid Co., Ltd.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Wu, Z.; Zhang, C.; Lei, B.; Zhao, J. Small Object Detection Based on Improved RetinaNet Model. *Comput. Simul.* **2023**, *40*, 181–189.
2. Zhao, Z.; Jiang, Z.; Li, Y.; Qi, Y.; Zhai, Y.; Zhao, W.; Zhang, K. Overview of visual defect detection of transmission line components. *Chin. J. Image Graph* **2021**, *26*, 2545–2560.
3. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
4. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
5. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
6. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
7. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
8. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
9. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525.
10. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
11. Jocher, G.; Changyu, L.; Hogan, A.; Yu, L.; Rai, P.; Sullivan, T. Ultralytics/Yolov5: Initial Release. (Version v1.0) [Z/OL]. 2020. Available online: <https://github.com/ultralytics/yolov5> (accessed on 20 June 2023).
12. Liu, S.; Qi, L.; Qin, H. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8759–8768.
13. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 11534–11542.
14. Zhang, Y.; Ren, W.; Zhang, Z.; Jia, Z.; Wang, L.; Tan, T. Focal and efficient IOU loss for accurate bounding box regression. *Neurocomputing* **2022**, *506*, 146–157. [[CrossRef](#)]

15. Wu, J.; Bai, L.; Dong, X.; Pan, S.; Jin, Z.; Fan, L.; Cheng, S. Transmission Line Small Target Defect Detection Method Based on Cascade R-CNN Algorithm. *Power Grids Clean Energy* **2022**, *38*, 19–27+36.
16. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
17. Huang, Q.; Dong, J.; Cheng, Y.; Zhu, Y. A Transmission Line Target Detection Method with Improved SSD Algorithm. *Electrician* **2021**, *282*, 51–55.
18. Zou, H.; Jiao, L.; Zhang, Z. An improved YOLO network for small target foreign body detection in transmission lines. *J. N. Inst. Technol.* **2022**, *20*, 7–14.
19. Li, J.; Liang, X.; Wei, Y. Perceptual generative adversarial networks for small object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1222–1230.
20. Liu, Y.; Wu, T.; Jia, X. Multi-scale target detection method for transmission lines based on feature pyramid algorithm. *Instr. User* **2019**, *26*, 15–18.
21. Li, L.; Chen, P.; Zhang, Y. Transmission line power devices and abnormal target detection based on improved CenterNet. *High-Volt Technol.* **2023**, 1–11. [[CrossRef](#)]
22. Su, X.; Zhang, M.; Chen, J.; Ding, Z.; Xu, H.; Bai, W. Transmission line tower target detection based on integrated YOLOv5 algorithm. *Rural Electr.* **2023**, 33–39. [[CrossRef](#)]
23. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
24. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 12993–13000.
25. Cao, S.; Deng, J.; Luo, J.; Li, Z.; Hu, J.; Peng, Z. Local Convergence Index-Based Infrared Small Target Detection against Complex Scenes. *Remote Sens.* **2023**, *15*, 1464. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.