

Article

Not peer-reviewed version

---

# Transmission Lines Small Target Detection Algorithm Research Based on YOLOv5

---

[Qiuyan Cheng](#) , [Guowu Yuan](#) , Dong Chen , Bangwu Xu , Enbang Chen , Ying Ren , [Hao Zhou](#) \*

Posted Date: 21 June 2023

doi: 10.20944/preprints202306.1543.v1

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



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

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

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>, Ying Ren<sup>2</sup> and Hao Zhou<sup>1,\*</sup>

<sup>1</sup> School of Information, Yunnan University, Yunnan 650504, China; chengqiuyan@itc.ynu.edu.cn(Q.C.); gwyuan@ynu.edu.cn(G.Y.); 12021215257@mail.ynu.edu.cn(D.C.); 852029986@qq.com(B.X.)

<sup>2</sup> Kunming Enersun Technology Co. LTD, Yunnan 650500, China; 39077057@qq.com(E.C.); 5865786@qq.com(Y.R.)

\* Correspondence: zhouhao@ynu.edu.cn

**Abstract:** The images captured by UAVs during inspection often contain **numerous** small targets related to transmission lines, which are critical and vulnerable elements 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 line 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, in order to address the challenges posed by the complex background and environmental interference that impact small targets, an ECA attention module is integrated into the network. This 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, which is 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, as they tend to aggregate. Experimental results demonstrate that the proposed P2E-YOLOv5 detection algorithm achieves an accuracy P 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

## 1. Introduction

In recent years, UAV technology has been widely used in the inspection of transmission lines. The images captured by 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 to define 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 small targets (assuming a default input image size of 640×640). The other approach is the absolute size definition. For instance, in COCO data, targets smaller than 32×32 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 but abundant 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 interfere 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 its 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, mainly represented by R-CNN<sup>[3]</sup>, Fast R-CNN<sup>[4]</sup>, Faster R-CNN<sup>[5]</sup> etc. Such algorithms usually generate target candidate regions, and then extract features through convolutional neural networks to achieve classification and localization prediction of target objects. In literature <sup>[6]</sup>, based on Faster R-CNN, a Feature Pyramid Network (FPN) was proposed to replace the original RPN and fuse high-level features rich in semantic information with lower-level features rich in positional information to generate feature maps of different scales. Then the target of the corresponding scale is predicted on the feature map of different scales.

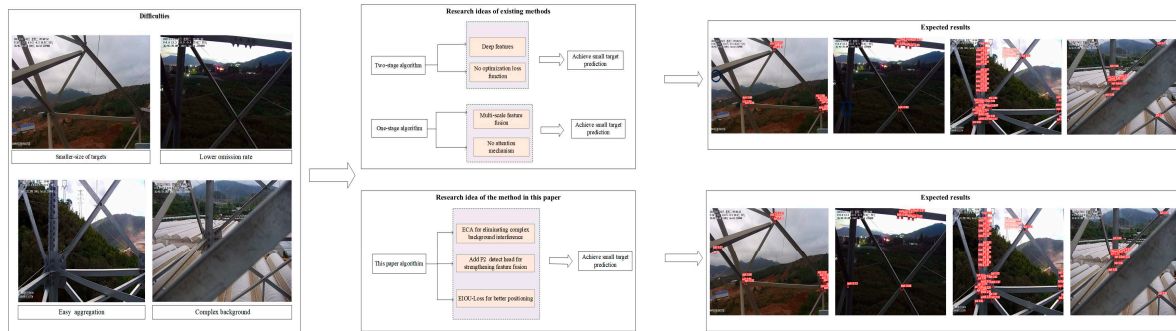
2) Single-stage detection algorithms, mainly represented by SSD <sup>[7]</sup>, YOLO <sup>[8-11]</sup> series, etc. Such algorithms can achieve target classification and positioning prediction by directly extracting features through convolutional neural networks. Redmon et al proposed YOLO <sup>[8]</sup> detection algorithm, which divides the image into  $S \times S$  grids and directly predicts the category probability and regression position information based on the surrounding box corresponding to each grid. This method does not generate candidate regions and improves the prediction speed. In the same year, Liu et al. <sup>[7]</sup> put forward the SSD algorithm, which draws on the idea of the YOLO algorithm and uses multi-scale learning on this basis 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<sup>[11]</sup>. It uses a deep residual Network to extract target features and uses a combination of feature pyramid network FPN and Perceptual Adversarial Network (PAN) <sup>[12]</sup> to realize the efficient fusion of rich low-level and high-level feature information. It realizes multi-scale learning and can effectively improve 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 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, the fusion of shallow and deep features in the multi-scale feature fusion is insufficient, leading to suboptimal accuracy in detecting 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 of 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 highly precise detection of small targets in transmission lines. The specific implementation of this method is as follows:

1. To tackle the challenge of complex background and susceptibility to environmental interference in small targets, we enhance the network's capability to focus on these targets and mitigate environmental disturbances by incorporating the ECA (Efficient Channel Attention) <sup>[13]</sup> attention module. This addition improves the network's attention towards small targets while simultaneously reducing the impact of environmental interference.
2. In response to the imaging characteristics of small size and lower-resolution targets, a high-resolution P2 detection head is incorporated into network to enhance the detection ability for small targets.

- Aiming at the characteristics that the small target features are not obvious and easy to aggregate, EIOU-Loss<sup>[14]</sup> is used as the regression loss function of the network to improve the accurate identification and positioning of small targets by the network.

In order to observe the characteristics of existing target detection algorithms and those proposed in this paper more concisely, a detailed research motivation diagram is drawn, as shown in Figure 1.



**Figure 1.** Among them, the one marked by the blue circle is the missed detection target. By comparing the two methods, it can be expected that the expected detection effect of the existing method is worse than that of the proposed method.

Moreover, 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 target and reduces the influence of environmental factors, making it highly suitable for small target detection in transmission lines. Furthermore, this algorithm can also achieve 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 difficult problem in the field of target detection, which has gradually become a research hotspot at home and abroad in recent years.

Wu et al.<sup>[15]</sup> proposed a method to detect small target defects of transmission lines based on the Cascade R-CNN algorithm. Specifically, ResNet101<sup>[16]</sup> network was used to extract target features, and a multi-hierarchy detector was used to identify and classify small targets of transmission lines. Experiments show that this algorithm can greatly improve the detection accuracy of small target defects in transmission lines. Huang et al.<sup>[17]</sup> and others proposed an improved SSD small target detection algorithm, PA-SSD, which integrated deconvolution fusion units into the PANet algorithm to fuse feature maps of different scales, strengthened multi-scale feature fusion, and formed a new feature pyramid model by replacing feature maps in the original SSD algorithm. Experiments show that the improved algorithm can extract feature maps with higher resolution, strengthen feature extraction for small targets, and significantly improve the detection accuracy of power components of transmission lines under the premise that the detection speed meets the detection performance.

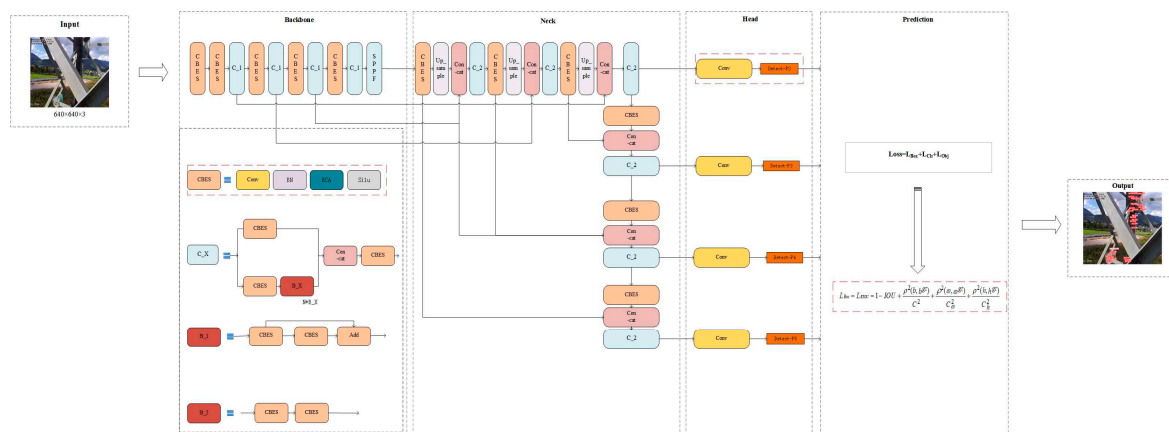
Zou et al.<sup>[18]</sup> proposed to effectively expand the foreign body data set of transmission lines through scene enhancement, Mixup, and noise simulation, thereby reducing the problem of fewer pictures in the data set and enriching the diversity of the data set, which is conducive to improving the detection performance of the detection network on small targets. Li et al.<sup>[19]</sup> proposed a perceptive GAN method specifically for small target detection, which learns high-resolution feature representations of small targets employing a generator and discriminator against each other. Liu et al.<sup>[20]</sup> proposed to replace the original VGG network in SSD structure with the ResNet network, thus enhancing the feature extraction capability of the network. Finally, FPN is added to the whole network structure to realize the information fusion of the upper and lower feature layers. Experiments show that compared with the original SSD algorithm, the accuracy of small and

medium-scale target detection in transmission line images is significantly improved. Li et al. [21] proposed an object detection algorithm based on improved CenterNet, which is to build a multi-channel feature enhancement structure and introduce the underlying details to improve the low detection accuracy caused by the use of a single feature by CenterNet. The research results show that the proposed algorithm has a good detection effect on power components and abnormal targets of transmission lines.

In summary, it can be found that the above research can't simultaneously take into account the detection accuracy and detection speed when carrying out small target detection of transmission lines, which is needed to achieve small target detection of transmission lines. Therefore, this paper intends to adopt the YOLOv5 detection model with good detection accuracy and fast detection speed as the basic model for improvement, to achieve higher detection accuracy while satisfying real-time detection.

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

In order to accurately and efficiently detect small targets in transmission lines, this paper proposes the use of the YOLOv5s network model with a simplified network architecture. The P2E-YOLOv5 network, as illustrated in Figure 2, demonstrates the basic structure of proposed approach.



**Figure 2.** P2E-YOLOv5 network architecture mainly consists of Input, Backbone (backbone network), Neck (neck network), Head (detection head), Prediction and Output. The red rectangular dotted line box marks the work done in this paper.

Aiming at the problems of the complex background of small targets of transmission lines and their vulnerability to environmental factors, this paper considers the method of adding an ECA attention mechanism to solve this problem. The CBES module is composed by incorporating an ECA attention module after the BN layer of the CBS module (the detailed structure is shown in Figure 5), which is widely distributed in the Backbone and Neck parts, which can promote the network to pay more attention to the small target area, reduce complex background interference, and enhance the feature extraction ability of the network for small targets; In view of the characteristics of small targets in transmission lines such as small proportion and low resolution, consider adding a P2 detection head in the Head part, which has a high resolution of  $160 \times 160$  pixels, enhances the effective fusion of rich low-level features and deep semantic information, and is generated by low-level and high-resolution feature maps, so it is very sensitive to the detection of small targets. In view of the characteristics of small targets of transmission lines that are not obvious and easy to aggregate, EIOU-Loss is considered as the loss function of network regression loss in the Prediction part, which can effectively improve the regression accuracy of the bounding box and enhance the accurate positioning and identification of small targets by the network.

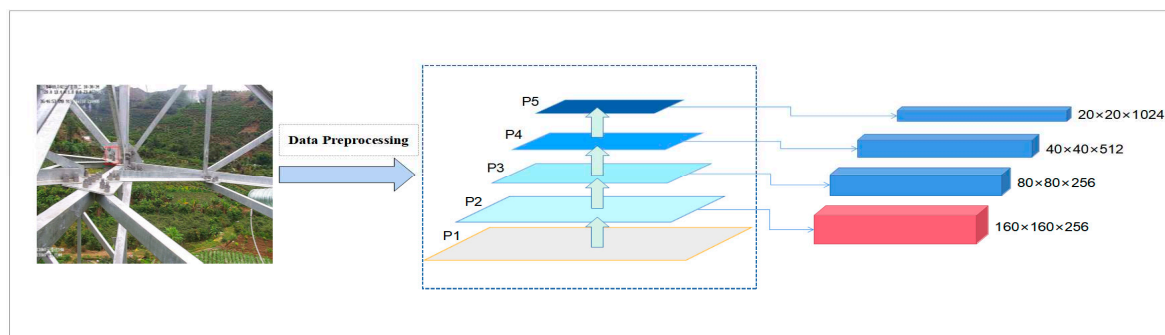
### 3.1. High resolution small target detection head

In order to solve the problems of small target size and low resolution of transmission line, this paper considers adding a high-resolution P2 detection head to the Head part of YOLOv5 to strengthen the feature fusion between shallow features of small targets and deep semantic information, so as to strengthen the feature extraction ability of small targets.

For the original YOLOv5 network model, the backbone network obtains 5-layer (P1, P2, P3, P4, P5) feature expression after 5 down-sampling operations, and P<sub>j</sub> indicates that the resolution is 1/2<sup>j</sup> of the input image. In the Neck part of the network, the top-down FPN and bottom-up PAN structures are used to achieve feature fusion at different scales, and then the object detection is carried out by the detection head drawn out on the three-level feature map P3, P4, and P5.

However, this paper is for the transmission line small target dataset for small target detection, and in fact, because the pixel scale of transmission line small targets in the image is often less than 32×32, such targets after multiple down-sampling, most of the detailed features and spatial information will be lost, even though the higher resolution P3 layer detection head can't detect these small targets.

In order to realize the detection of the above small targets, a new detection head is introduced in the P2 layer feature map of the YOLOv5 network model, as shown in Figure 3.



**Figure 3.** Based on the three detection heads extracted from the original P3, P4 and P5 layer feature map, detection head of a new 160\*160 pixels high-resolution is extracted from the P2 layer feature map.

First of all, the detection head has a high resolution of 160×160 pixels in size, can detect tiny targets with a scale of 8×8 pixels, and is very sensitive to smaller targets. The new high-resolution P2 detection head is equivalent to only two down-sampling operations on the backbone network, and extremely rich shallow feature information of small targets is obtained.

Secondly, by merging the top-down features of the P2 layer with the same-scale features of the backbone network in the Neck part, shallow features, and deep semantic features can be fused. In this way, the output features obtained are the result of an effective fusion of multiple input features.

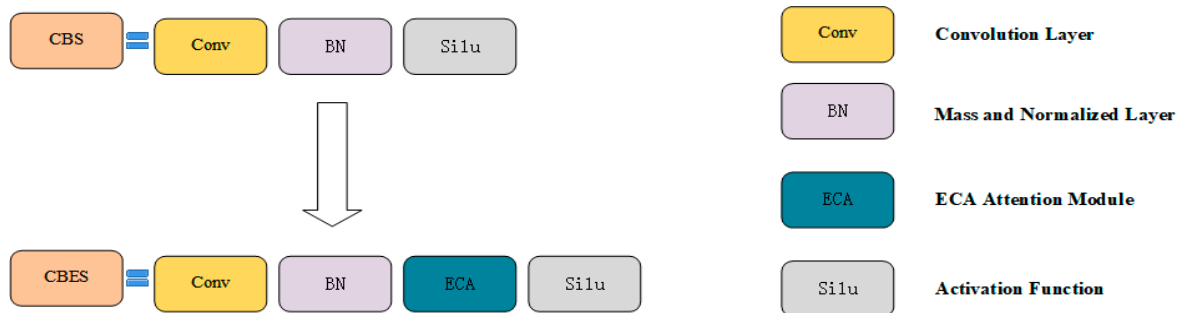
Finally, the P2 layer detection head combined with the original three detection heads has a good effect on alleviating the negative impact of scale variance. The P2 layer detection head is generated from low-level, high-resolution feature maps, which are rich in small target detail features and semantic information.

Therefore, it will be very sensitive to the detection of small targets in transmission lines, even if the addition of this detection head will make the calculation amount and memory of the model larger, but its improvement in the detection accuracy of small targets is very considerable.

### 3.2. ECA attention mechanism

For the small target images of transmission lines with complex environmental backgrounds, such as representative insulators and bolts, there are problems such as a lack of obvious features and extremely susceptible to interference by environmental factors, which makes it difficult for the

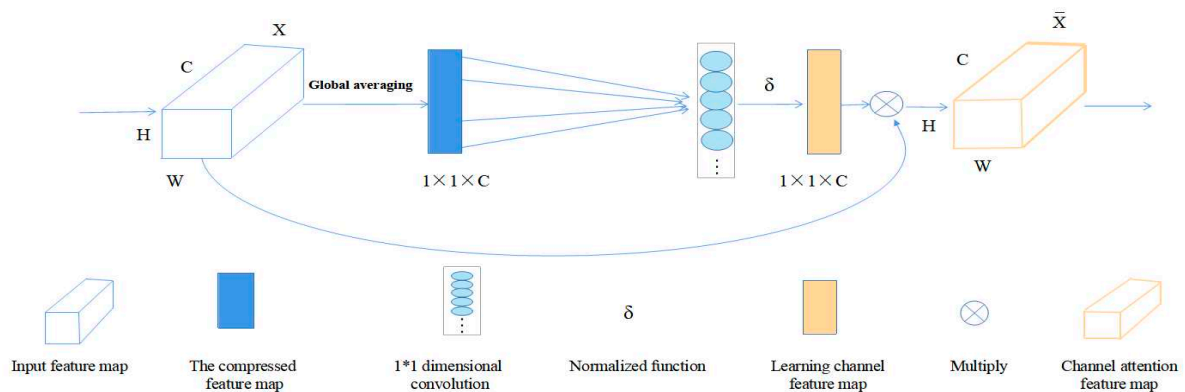
detection model to accurately detect them. Therefore, this paper adopts the ECA attention mechanism to solve the above problems, and the additional form is shown in Figure 4.



**Figure 4.** The ECA attention mechanism is added after the BN layer of the CBS module.

By adding the ECA attention module after the BN layer of the CBS module, the interference of environmental factors on small targets can be greatly reduced, the small targets that are blocked can be effectively identified, and the attention of the model to the small target area can be strengthened, and the CBES module composed of the ECA attention module can be added to the backbone network and neck network, which can strengthen the extraction ability of the network model on the features of small targets and improve its detection performance.

ECA is a lightweight attention module that can be directly integrated into the CNN architecture, which uses a local cross-channel interaction strategy without dimensionality reduction, which effectively avoids the influence of dimensionality reduction on the channel attention learning effect. Proper cross-channel interaction can significantly reduce model complexity while maintaining performance. Its structure is shown in Figure 5,



**Figure 5.** Specific composition of ECA attention module.

The ECA Attention module is described in detail as follows:

- (1) First input H\*W\*C dimension feature graph.
- (2) The input feature map is compressed for spatial features: in the spatial dimension, the feature map of 1\*1\*C is obtained by using global average pooling.
- (3) Learn channel features for the compressed feature graph: learn the importance of different channels through 1\*1 convolution, and the output dimension is still 1\*1\*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 adaptively, 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  $1*1*C$  and the original input feature map  $H*W*C$  are multiplied channel by channel to generate a feature map with channel attention.

### 3.3. Optimize the loss function

In order to solve the problem that the characteristics of small targets on transmission lines are not obvious and easy to gather, this paper considers using EIOU-Loss as the loss function of network regression loss to improve the regression accuracy of boundary frame and enhance the accurate identification and positioning of small targets by the network.

For small target images of transmission lines captured by UAV, typical targets such as insulators and bolts account for a very small proportion of the whole image, with low resolution and little visualization information, making it difficult to extract clear and obvious detailed features. These characteristics make it difficult for detection models to accurately locate and identify them. In addition, small targets have a high probability of aggregation. When small targets are dense, the small targets adjacent to the aggregation region will converge and integrate a point after multiple down-sampling into the deep feature map, making the detection model unable to distinguish. At the same time, when small targets are too close, a large number of correct prediction boundary frames may be filtered out due to the non-maximum suppression operation of post-processing. So as to lead to the detection of the case of missing.

The original YOLOv5 network adopted CIOU-Loss<sup>[22]</sup> as the loss function of network regression loss. CIOU-Loss took into account the overlap area, center point distance, and aspect ratio of boundary frame regression, but the aspect ratio of boundary frame regression was only for the difference in the aspect ratio of boundary frame, rather than the real difference between width and height respectively and their confidence. Moreover, when the aspect ratio of the prediction frame changes proportionally, the aspect ratio of the boundary frame regression penalty will lose its function, and it will become difficult to locate and recognize small targets. Therefore, in this paper, the EIOU-Loss function is used as the regression loss function of the P2E-YOLOv5 model, which solves the problem of fuzzy definition of aspect ratio and proportional change of aspect ratio to punish failure in CIOU-Loss, which can effectively improve the detection model's accurate positioning of small targets such as insulators and bolts. The convergence effect of the model is enhanced and the omission rate of the model to these small targets is reduced effectively.

The EIOU-Loss function adopted in this paper includes three parts: Overlap loss  $L_{IOU}$ , center distance loss  $L_{dis}$ , and width and height loss  $L_{asp}$ . The first two parts continue the advantages of the method in CIOU and width and height loss are to divide the loss term of aspect ratio into the difference between the predicted width and height respectively and the minimum width and height of the external frame on the basis of the penalty term of CIOU, to accelerate the model convergence and improve the regression accuracy. The calculation formula 2 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 ratio between the intersection and union of the prediction box, and ground truth box,  $b$  and  $b^{gt}$  represents the center point of the prediction box and ground truth box,  $q(b, b^{gt})$  represents the Euclidean distance between the center point of the prediction frame and the ground truth frame,  $C$  represents the diagonal distance of the smallest outside rectangle of prediction box and ground truth box,  $\omega^{gt}$  and  $h^{gt}$  are respectively the length and width of ground truth box.  $\omega$  and  $h$  are the length and width of the prediction frame respectively, while  $C_\omega$  and  $C_h$  are the width and height of the minimum external frame covering the ground truth frame and prediction frame.

## 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 Pytorch1.7.1.

#### 4.1. Evaluation indicators and data set preparation

Accuracy P(Precision), Recall R(Recall), Average Precision AP (Average Precision), mean average precision mAP (mean Average Precision), and Frames Per second were used in this paper. FPS evaluates the model from these aspects. The calculation expressions of P, R, AP, and mAP are shown in Formula 3, 4, and 5:

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (3)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (4)$$

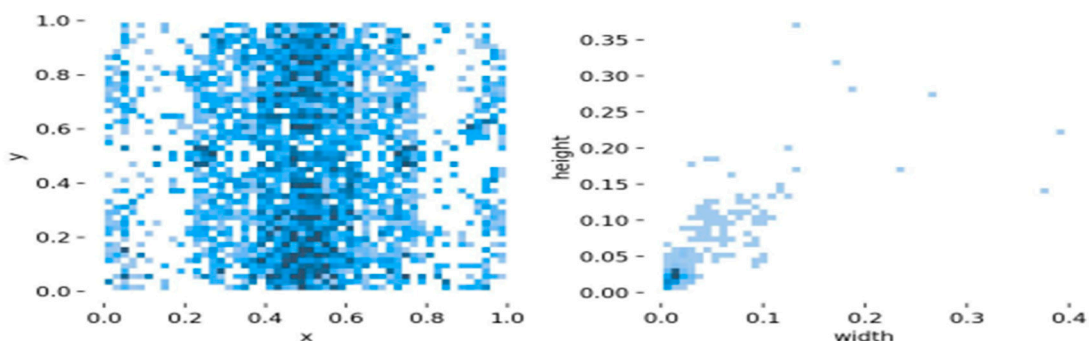
$$AP = \int_0^1 P(R)dR, mAP = (\sum_i^N AP_i) / N \quad (5)$$

where  $N_{TP}$ ,  $N_{FP}$ , and  $N_{FN}$  are respectively the number of samples predicted correctly, the number of samples predicted incorrectly and the number of samples not recognized, and  $N$  is the number of categories set by the model.

The original data used in this experiment comes from the small target data set of transmission lines provided by a power grid company in Yunnan, and the image contains abundant small targets of transmission lines such as insulators and bolts. By performing random horizontal flip, rotation transformation, brightness transformation, and other data expansion operations on 1076 images of the original data set, a new data set containing 3200 images was obtained. The new data set was annotated by image labeling software labelImg and divided into the training set, verification set, and test set according to the ratio of 8:1:1.

#### 4.2. Experiment

The small target data set of transmission lines adopted in this experiment after data processing was analyzed, as shown in Figure 6.



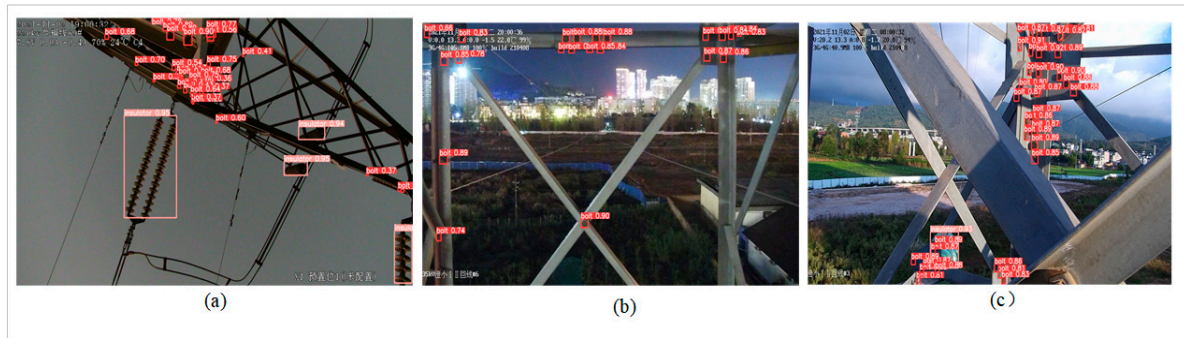
**Figure 6.** Data analysis of transmission line small target data set.

In Figure 7, the figure on the left shows the distribution of the center coordinates of the small target object in the small target data set of the transmission line. It can be seen that the central coordinates of the small target object are centrally distributed in the range (0.4~0.6,0.0~1.0). The figure on the right shows the height and width distribution of small target objects in the data set relative to the whole picture. It can be seen that small target objects are very small compared to the whole picture. In the figure on the right, most small target objects account for far less than 0.12% of the area

of the whole picture (the definition of the relative size of small targets is explained in detail in the introduction section). It indicates that most of the targets in this data set are small targets.

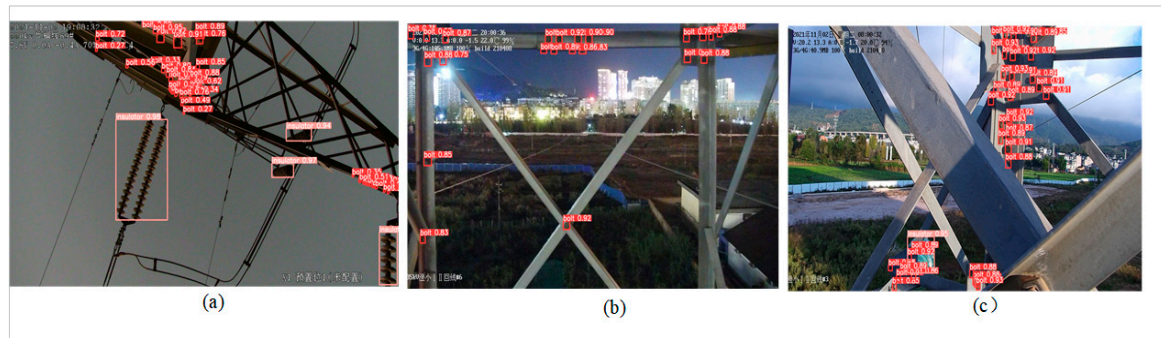
The training experiment was carried out on the small target data set of transmission lines, and the original YOLOv5 training results were compared with the improved P2E-YOLOv5 training results:

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



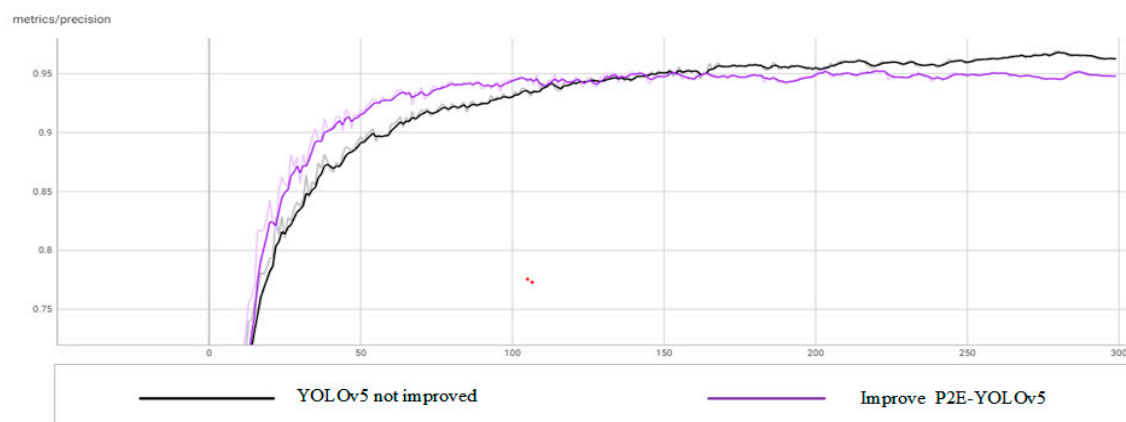
**Figure 8.** Test results before improvement. (a-c) are three pictures of random test before model improvement.

The improved model test results are shown in Figure 9:

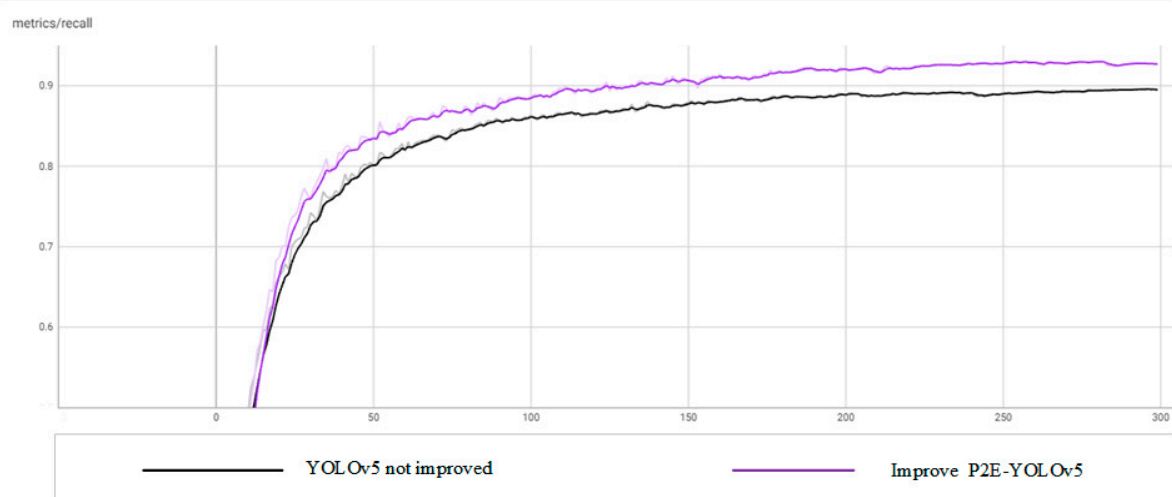


**Figure 9.** Improved test results. (a-c) are three pictures of random test after model improvement.

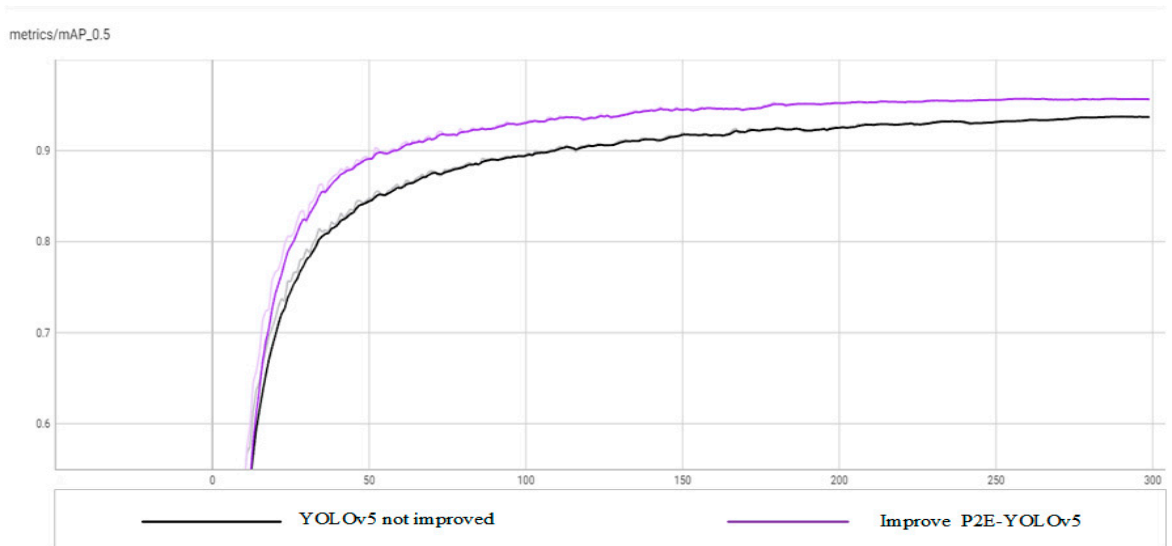
Model performance comparison results are shown in Figures 10–13:



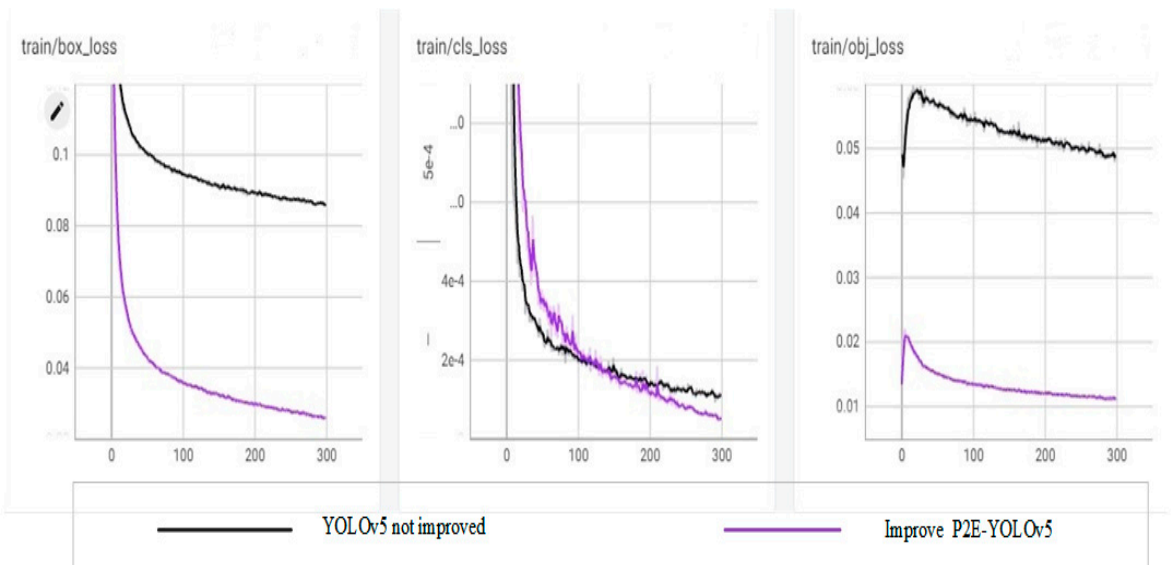
**Figure 10.** Comparison results of accuracy.



**Figure 11.** Comparison results of recall rates.



**Figure 12.** Average precision mAP comparison results.



**Figure 13.** Comparative results of training losses.

From the above comparison results, it can be seen that compared with the original YOLOv5, the improved P2E-YOLOv5 has significantly improved the performance of accuracy, recall rate, average accuracy, and loss index, and significantly improved the detection ability of objects with low resolution and small size in images.

#### 4.3. Comparative experiment

In order to prove the effectiveness of this model, based on the existing small target data set of transmission lines, the paper chooses Faster-RCNN, RetinaNet, SSD, YOLOv3\_spp, YOLOX, and YOLOv5 as the comparison network model. The experimental results are shown in Table 1.

**Table 1.** Comparison of experimental results of different network models. In addition, mAP\_0.5 represents the mean precision when the intersection ratio is 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

It can be seen from Table 1 that for the mAP with mean average accuracy that can reflect model detection performance, compared with the speed-RCNN method, the optimized mAP of P2E-YOLOv5 is improved by 9.8%. Compared with the optimized RetinaNet method, the mAP of P2E-YOLOv5 increased by 18.7%; Compared with the SSD method, mAP of P2E-YOLOv5 after optimization is significantly improved by 43.5%. Compared to the YOLOv3\_spp method, the optimized P2E-YOLOv5 mAP is 17.2% higher. Compared with the YOLOX method, the mAP of P2E-YOLOv5 improved by 6.6% after optimization. Compared with the original YOLOv5 method, the overall mAP of P2E-YOLOv5 improved by 3.3% after optimization. It is obvious that the P2E-YOLOv5 detection method proposed in this paper is far superior to other detection algorithms in terms of detection accuracy.

In addition, in terms of the detection speed of the model, compared with other controls, the optimized P2E-YOLOv5 reached 87.719 frames, second only to the detection speed of the original YOLOv5. The reason why the detection speed of P2E-YOLOv5 is lower than that of the original YOLOv5 is that the newly added P2 detection header will increase a certain amount of computation, which makes the reasoning time of the model longer. Therefore, the reasoning detection speed decreases somewhat. However, compared with other detection methods, P2E-YOLOv5 still has a great advantage, which can meet the requirements of real-time detection.

In summary, compared with other target detection algorithms, all indexes of the P2E-YOLOv5 algorithm proposed in this paper have been greatly improved, effectively solving the problem of obtaining high detection accuracy in small target detection of transmission lines and meeting the requirement of real-time detection.

#### 4.4. Ablation experiment

Based on the preprocessed transmission line small target data set, the enhancement effect of each new or improved module on the overall model was explored through the ablation experiment. Based on the original YOLOv5, each module was added successively for experiments until the final model P2E-YOLOv5 was obtained. The experimental results are shown in Table 2.

**Table 2.** 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

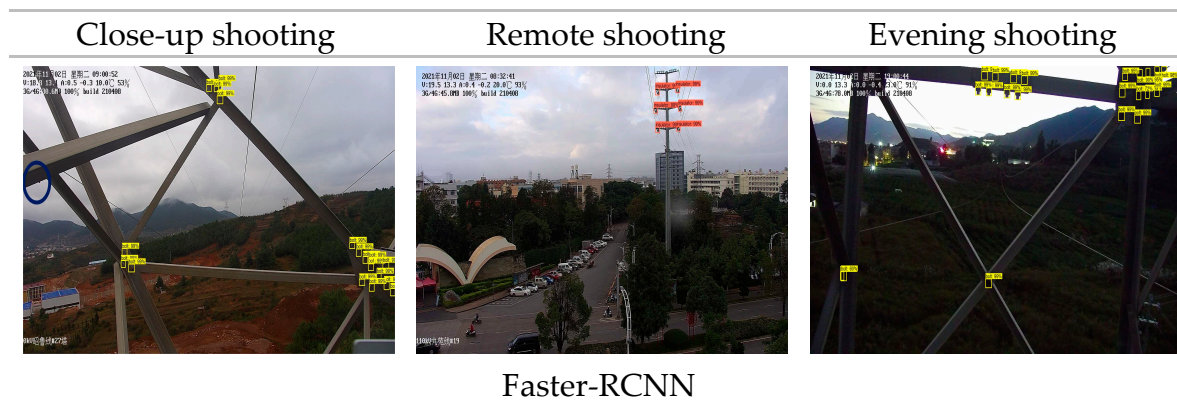
As can be seen from Table 2, in terms of detection performance improvement, when the small target detection head is added, the accuracy P of the model is slightly decreased, but the recall rate R and the mean average accuracy mAP are increased by 3.5% and 2.2%, respectively. It can be seen that the addition of a P2 layer small target detection head greatly improves the model, indicating that it can effectively enhance the detection ability of the network for small targets. Next, the ECA attention module was added on the basis of adding small target detection heads. It could be clearly observed that although the mean average accuracy of the model did not change significantly, its accuracy P and recall rate R increased by 0.4% and 0.3% respectively, indicating that it could effectively strengthen the network's attention to small target objects and improve the feature extraction ability of the model for small targets. Finally, on the basis of adding a small target detection head and ECA attention module, the loss function EIOU-Loss was added. Although the recall rate R of the model decreased, its accuracy P and mean average accuracy mAP were both improved, indicating that it could effectively improve the detection accuracy of the model for small targets.

In addition, in the aspect of detection aging, the inference detection speed of the model decreased due to the increase in calculation amount caused by the amplification of the detection layer scale. However, after the addition of the ECA attention module and EIOU-Loss function, it was obviously observed that the detection speed increased.

Through the above data analysis results, it can be proved that it is consistent with the theoretical analysis results, which again verifies the rationality and effectiveness of the algorithm in this paper while satisfying the real-time detection and improving the detection accuracy.

#### 4.5. Detection Effect

In order to verify the effectiveness of the P2E-YOLOv5 model for small target detection of transmission lines, the expanded small target data set training weight number is used to detect the image of the test set, and three transmission line images in different scenes are randomly selected for detection. The test results are shown in Figure 14.





RetinaNet



SSD



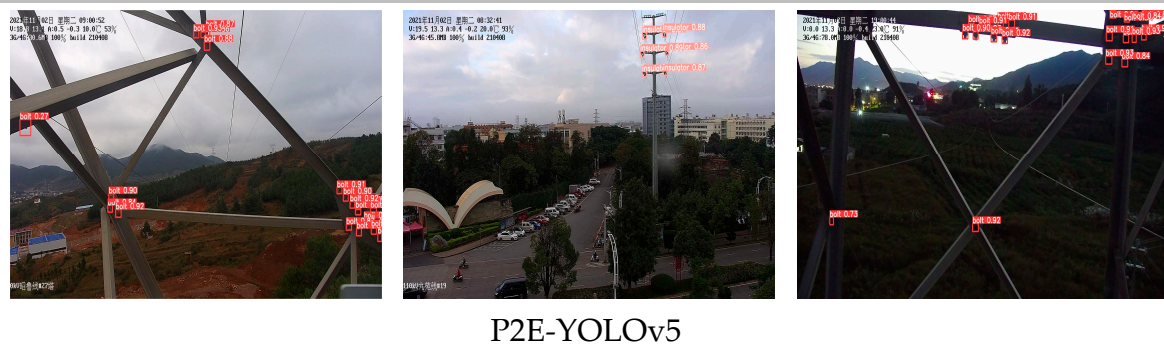
YOLOv3\_spp



YOLOX



YOLOv5



**Figure 14.** This figure shows in detail the detection effect of several detection algorithms such as Faster-RCNN, RetinaNet, SSD, YOLOv3\_spp, YOLOX, YOLOv5, and P2E-YOLOv5 on small targets in transmission path images in different scenes.

In Figure 14, test results of Faster-RCNN, RetinaNet, SSD, YOLOv3\_spp, YOLOX, YOLOv5, and several detection algorithms proposed in this paper for small targets in different transmission path images, which, marked by a rectangle box the detected small target. The blue circles indicate small missed targets.

It can be observed that Faster-RCNN and YOLOX detection algorithms can accurately detect most small target objects with high detection accuracy, but for some small target objects with shallow features, there are some missing detection cases. Although the detection algorithms such as RetinaNet, SSD, and YOLOv3\_spp can detect most small targets, the detection accuracy is low and there are many omissions. YOLOv5 detection algorithm has a high detection accuracy and can detect most small targets, but there are still some small targets missing detection.

By comparing the above 6 detection algorithms, the P2E-YOLOv5 detection algorithm can not only accurately identify small targets in different scenes, but also has higher detection accuracy and the lowest missing rate. The reasons are as follows: First, the added ECA attention module effectively improves the network's attention to small targets and greatly reduces the interference of environmental factors; Secondly, the newly added high-resolution P2 detection header greatly improves the detection ability of the network to small targets. Finally, EIOU-Loss is used as the regression loss function of the network, which makes the identification and location of small targets more accurately. Therefore, the P2E-YOLOv5 detection algorithm achieves a better detection effect than the other 6 detection algorithms on a small target detection data set of transmission lines.

## 5. Conclusion

In this paper, a P2E-YOLOv5 detection method is proposed to address the challenge of accurately and efficiently detecting small targets, such as insulators and bolts, which are important components of transmission lines. By incorporating the ECA attention mechanism into the network and introducing a high-resolution detection head that is more sensitivity to small targets, and optimizing the network's regression loss function using EIOU-Loss, our proposed method significantly improves the detection efficiency and capability of small targets in complex backgrounds. We evaluate the performance of our algorithm on a dataset specifically designed for small target detection in transmission lines. Compared to Faster-RCNN, RetinaNet, SSD, YOLOv3\_spp, YOLOX and YOLOv5 methods, our method achieves remarkable improvements in detection accuracy of 9.8%, 18.7%, 43.5%, 17.2%, 6.6% and 3.3%, respectively. Although the increased memory usage of the newly added high-score detection head slightly affects the computational reasoning speed, our method still outperforms Faster-RCNN, RetinaNet, SSD, YOLOv3\_spp and YOLOX methods. Considering the trade-off between detection accuracy and speed, our P2E-YOLOv5 detection method is the optimal choice. Future work aims to reduce the computational cost associated with the P2 detection head while maintaining comparable detection accuracy and further improving the network detection speed.

**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. and Y.R.; data curation, H.Z.; 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 Major science and technology project of Yunnan Province, grant number 202202AD080004.

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Restrictions apply to the availability of these data. Data was 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[J/OL]. *Computer Simulation*:1-10[2022-11-29].
2. Zhao, Z.; Jiang, Z.; Li, Y.; Qi, Y.; Zhai, Y.; Zhao, W.; Zhang, K. Overview of visual defect detection of transmission line components[J]. *Chinese Journal of Image and Graphics*,2021,26(11):2545-2560.
3. Ross, B.; Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation[J]. *CoRR*,2013, abs/1311.2524.
4. Girshick, R. Fast R-CNN[C]. *IEEE International Conference on Computer Vision*. Boston. MA. USA. 2015. 1440-1448.
5. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39 (6):1137-1149.
6. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection[C]. *IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu. HI. USA. 2017. 936-944.
7. Liu, W.; Anguelov, D.; Erhan, D. SSD: Single shot multibox detector[C]. *European Conference on Computer Vision*. Amsterdam. Netherlands. 2016. 21-37.
8. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection[C]. *IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas. NV. USA. 2016. 779-788.
9. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger[C]. *IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu. HI. USA. 2017. 6517-6525.
10. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal speed and accuracy of object detection[J]. *arXiv preprint*, arXiv:2004.10934, 2020.
11. Jocher, G. YOLOv5[DB/OL]. <https://github.com/ultralytics/yolov5>, 2020.
12. Liu, S.; Qi, L.; Qin, H. Path aggregation network for instance segmentation[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018: 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[J]. *CoRR*,2019, abs/1910.03151.
14. Zhang, Y.; Ren, W.; Zhang, Z.; Jia, Z.; Wang, L.; Tan, T. Focal and efficient IOU loss for accurate bounding box regression[J]. *Neurocomputing*,2022,506.
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,2022,38(04):19-27+36.
16. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition[J]. *CoRR*,2015, abs/1512.03385.
17. Huang, Q.; Dong, J.; Cheng, Y.; Zhu, Y. A Transmission Line Target Detection Method with Improved SSD Algorithm [J]. *Electrician*,2021, No.282(06):51-55.
18. Zou, H.; Jiao, L.; Zhang, Z. An improved YOLO network for small target foreign body detection in transmission lines [J]. *Journal of Nanjing Institute of Technology (Natural Science Edition)*,2022,20(03):7-14. DOI: 10.13960/j.issn.1672-2558.2022.03.002.

19. LI, J.; LIANG, X.; WEI, Y. Perceptual generative adversarial networks for small object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE,2017:1222-1230.
20. Liu, Y.; Wu, T.; Jia, X. Multi-scale target detection method for transmission lines based on feature pyramid algorithm [J]. Instrument user,2019,26(01):15-18.
21. Li, L.; Chen, P.; Zhang, Y. Transmission line power devices and abnormal target detection based on improved CenterNet [J/OL]. High-Voltage Technology:1-11[2023-03-30].
22. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression[J]. Proceedings of the AAAI Conference on Artificial Intelligence,2020,34(07).

**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.