



# Article Transmission Line Fault Detection and Classification Based on Improved YOLOv8s

Hao Qiang <sup>1,2</sup>, Zixin Tao <sup>1</sup>, Bo Ye <sup>1</sup>, Ruxue Yang <sup>1</sup> and Weiyue Xu <sup>1,\*</sup>

- <sup>1</sup> School of Mechanical Engineering and Rail Transit, Changzhou University, Changzhou 213164, China; qhao@cczu.edu.cn (H.Q.); s22050858027@smail.cczu.edu.cn (Z.T.); 20085800048@smail.cczu.edu.cn (B.Y.); s22050858030@smail.cczu.edu.cn (R.Y.)
- <sup>2</sup> Jiangsu Province Engineering Research Center of High-Level Energy and Power Equipment, Changzhou University, Changzhou 213164, China
- \* Correspondence: wyxu@cczu.edu.cn

Abstract: Transmission lines are an important component of the power grid, while complex natural conditions can cause fault and delayed maintenance, which makes it quite important to locate and collect the fault parts efficiently. The current unmanned aerial vehicle (UAV) inspection on transmission lines makes up for these problems to some extent. However, the complex background information contained in the images collected by power inspection and the existing deep learning methods are mostly highly sensitive to complex backgrounds, making the detection of multi-scale targets more difficult. Therefore, this article proposes an improved transmission line fault detection method based on YOLOv8s. The model not only detects defects in the insulators of power transmission lines but also adds the identification of birds' nests, which makes the power inspection more comprehensive in detecting faults. This article uses Triplet Attention (TA) and an improved Bidirectional Feature Pyramid Network (BiFPN) to enhance the ability to extract discriminative features, enabling higher semantic information to be obtained after cross-layer fusion. Then, we introduce Wise-IoU (WIoU), a monotonic focus mechanism for cross-entropy, which enables the model to focus on difficult examples and improve the bounding box loss and classification loss. After deploying the improved method in the Win10 operating system and detecting insulator flashover, insulator broken, and nest faults, this article achieves a Precision of 92.1%, a Recall of 88.4%, and an mAP of 92.4%. Finally, we conclude that in complex background images, this method can not only detect insulator defects but also identify power tower birds' nests.

Keywords: transmission line; fault detection; improved YOLOv8s; BiFPN

# 1. Introduction

Transmission line fault detection [1] is a key technology to ensure power supply reliability. In recent years, with the continuous development of the power industry, the proliferation of power grid equipment being constructed in challenging environments is on a constant rise every day, which has put forward higher requirements for the safety and maintenance of power system equipment. At present, the image data collected by unmanned aerial vehicle (UAV) patrol inspection [2] has been imported into the offline server. Its intelligence and operation autonomy speed up the efficiency of UAV uploading the collected fault information to the automatic master station system so as to better realize the defect location of the fault section [3].

With the continuous expansion of the power grid scale, transmission lines are prone to failure under the influence of external environmental factors [4]. Traditional transmission line operation faults include broken insulators, bird nests at the top of the power tower, and pollution flashover faults [5]. As an important component to ensure the normal operation of transmission lines, the insulator also has the highest probability of failure in transmission lines. Insulators will be affected by ice damage, wind damage, lightning, and other adverse



**Citation:** Qiang, H.; Tao, Z.; Ye, B.; Yang, R.; Xu, W. Transmission Line Fault Detection and Classification Based on Improved YOLOv8s. *Electronics* **2023**, *12*, 4537. https:// doi.org/10.3390/electronics12214537

Academic Editors: Azeddine Kaddouri and Nouha Bouchiba

Received: 21 September 2023 Revised: 31 October 2023 Accepted: 2 November 2023 Published: 4 November 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/). factors when exposed to complex environments for a long time, resulting in varying degrees of failure. Serious failures may affect the operation of the entire power system, leading to large-scale blackouts and resulting in huge economic losses [6]. Distance-based Protection is a traditional method that involves measuring impedance and reactance values to estimate the fault location. Although it is widely used, it can be less accurate for complex network configurations and may fail in the presence of high-impedance faults. Wavelet Transform Analysis employs wavelet analysis to examine the voltage and current signals on the transmission line. Although it can effectively detect various types of faults, it may be computationally intensive and require precise synchronization of data. Traditional fault classification methods based on mathematical modeling and signal processing typically require manual or robotic detection, which not only consumes a large number of human resources but also can easily lead to missed detection. Moreover, these methods rely on assumptions about linear and stationary systems, so they may also have problems with nonlinear and non-stationary signals [7].

At present, the detection of transmission line faults is mainly achieved through drone aerial photography. When compared to manual inspection, UAV inspections offer several advantages, including low cost, enhanced efficiency, and increased safety measures. However, the images collected by UAV patrol inspection usually include complicated backgrounds with the presence of forests, grasslands, rivers, farmland, etc., which causes detection uncertainty. In addition, the problems of shooting distance and shooting angle during patrol inspection will cause the collected target defects to be small and sheltered. Many researchers have begun to identify and locate defects [8,9] in drone aerial images to further improve the quality of transmission line detection. Machine Learning and deep learning methods [10] have greatly assisted in the detection, classification, and localization of defects in power systems. Currently, the mainstream deep learning-based object detection algorithms can be divided into two categories: two-stage algorithms based on candidate regions and one-stage algorithms based on regression. The former has a complex structure but has a high accuracy rate. While the latter has a simple structure and fast detection speed, its accuracy is relatively low. This algorithm directly performs regression on the input image through a convolutional neural network and then outputs the final result.

He et al. [11] once conducted research on the detection of insulators missing in aerial images based on deep learning. They first used Faster R-CNN to identify the transmission line insulator and then used the CNN algorithm to detect the defect of the missing insulator after obtaining satisfactory identification results. The accuracy of detecting missing defects in the article reached 86%, but it only focused on whether the insulator had defects and did not position the defects accurately. However, Lei et al. [12] proposed a deep convolutional neural network method based on the Faster R-CNN method to locate broken insulators and bird nests. This article detected two types of faults and converted the problem of target classification into the problem of target detection and identification. The detection performance was good, but the insulator defects were still not accurately located, and the detection time of a single picture was long (201 ms). Due to the long detection time and large model of the two-stage algorithm, the one-stage algorithm has more advantages in transmission line fault detection. Liu et al. [13] proposed an improved YOLOv3 network to detect insulators under different background interference. By adding dense blocks to optimize feature extraction and optimizing anchor frames through the clustering algorithm, the accuracy of insulator identification under complex backgrounds was enhanced. Although the article carefully divided the insulator detection results under different scenarios, it was limited to the identification of insulators and failed to divide the defective insulators and normal insulators. Wu et al. [14] proposed a lightweight YOLOv3 insulator defect detection method using a k-means++ algorithm based on Euclidean Distance to improve the stability of generating a priori frame. In order to improve the detection speed of the model, this article applied Crop-MobileNet to reduce the time cost, but the mAP was only 84%, which would make it more difficult to locate the fault in a complex background and easy to cause missed detection. Wu [15] et al. proposed an improved YOLOv5 algorithm for detecting

multiple defects of insulators to solve the problems of low detection accuracy and slow detection speed. In this article, GhostNet and SimAM modules were introduced to improve the accuracy and speed of the network for insulator flashover and damage, and the map reached 87.8%. However, my research not only detected two kinds of insulator defects but also added the birds' nest defect to study the fault detection accuracy under complex backgrounds, and the final mAP also reached a good percentage.

With the deepening of the deep learning algorithm, more attention mechanisms, such as Convolutional Block Attention Module (CBAM), Efficient Channel Attention (ECA), and so on, are integrated, which helps the neural network to dynamically search the regions with discriminant characteristics from the perspective of multi-channel attention fusion and cross element attention, so as to improve the detection accuracy. The current YOLO series has been updated to YOLOv8 [16], which provides a brand new state of the art (SOTA) model, replacing the C3 structure of YOLOv5 with a more gradient-flow-rich C2f structure and adjusting the number of channels for different scale models. YOLOv8 has been applied to different research objects, among which Chen et al. [17] performed better detection of aircraft targets in Synthetic Aperture Radar (SAR) images, eliminating the large target detection layer of YOLOv8 and introducing deformable convolution, thereby achieving a balance between model complexity and detection accuracy. Gao et al. [18] used a deformable convolutional improved backbone network and occlusion perception attention mechanism on the YOLOv8 model to detect dense pedestrians, improving the accuracy of small object detection.

To solve the problems, the deep learning model may be highly sensitive to complex backgrounds and would probably affect the detection of multi-scale targets in fault detection. This article proposes an improved algorithm based on the YOLOv8s network. By building upon the existing advanced single-stage target detector YOLOv8s, this method introduces several improvements to the network structure and loss function, enabling it to effectively address more intricate transmission line fault detection tasks. The main contributions of this article are as follows:

(1) Since the two defects on the insulator belong to different subcategories under the same category, an improved Mosaic image enhancement technique is adopted to enrich the image features and alleviate the accuracy decline caused by overfitting;

(2) To promote the performance of feature extraction during model training, an improved Triplet Attention (TA) is used to enhance the network's ability to calculate attention weights and apply them to key features. In addition, to reduce the generation of redundant parameters, we propose that a lighter Light Conv structure be substituted for a portion of the convolutional structure in YOLOv8s;

(3) In order to generate higher semantic information after fusion, two cross-layer connections are established in the feature pyramid network (FPN) to fuse the  $80 \times 80$  and  $40 \times 40$  feature maps across layers. A comparative analysis of four distinct fusion methods has determined that the "bifpn" exhibits the most optimal fusion technique;

(4) WIoU v2, a monotonic focusing mechanism for cross-entropy, is introduced to enable the model to focus on difficult examples and better convergence of boundary loss and classification loss, which improves the detection accuracy of the YOLOv8s model and the overall performance of it;

(5) The experimental results show that the improved YOLOv8s has a detection Precision of 92.1%, a Recall of 88.4%, and a mAP of 92.4%. For class-specific metrics, the mAP for pollution–flashover, broken, and nests are 87.6%, 91.4%, and 99.5%, respectively.

The rest of this article is organized as follows: Section 2 describes the constitution of the improved YOLOv8s detection network. Section 3 presents the relative experiments and analysis. Finally, Section 4 discusses and summarizes the whole article.

#### 2. Model Construction Based on Improved YOLOv8s

This article proposes an improved YOLOv8s [16] algorithm network structure, as shown in Figure 1. The model uses YOLOv8s as a framework, enriches the dataset with

improved Mosaic [19] enhancement techniques, and adds Triplet Attention [20] to the backbone network to assist in feature extraction for small targets. To improve the efficiency of feature fusion, two cross-layer connected Bidirectional Feature Pyramid Network (BiFPN) [21] networks are used, and deformable convolution is introduced to reduce the computational complexity. Finally, a combination of Distribution Focal Loss (DFL) [22] and Wise-IoU (WIoU) [23] is used as the regression loss.



Figure 1. Improved YOLOv8s network structure.

# 2.1. Improved Mosaic Data Enhancement

The YOLOv8 algorithm uses Mosaic [19] data enhancement to enrich the data set during the data preprocessing phase. The process of the algorithm combines four images through reversal and zooming operations and gamut variations to produce a Mosaic image. The Mosaic data enhancement of the four images is shown in Figure 2a. This article uses Mosaic data enhancement of nine images to increase the size of each batch of images for training, resulting in higher training efficiency. The mosaic enhancement of nine random images is shown in Figure 2b, and the green box is the annotation box. Moreover, the dataenhanced image is closer to the detection size of small targets, increasing the complexity of the detected image and preventing overfitting. YOLOv8 introduces the operation of closing Mosaic in the last 10 epochs proposed in YOLOX [24], which effectively improves detection accuracy.



(a) Four images moasic

(b) Nine images moasic

Figure 2. Comparison of the enhancement effect of the improved mosaic data.

#### 2.2. Improved Backbone Network

Feature extraction is a layer-by-layer process of mining image data. Due to its fixed convolution kernel and small receptive field, standard convolution operation may result in missed detection in multi-class transmission line fault detection. This article adds a Channel Attention to C2f [16], which can focus on more effective features. We adopt the Triplet Attention [20] to better extract the features of small targets. Finally, a lighter-weight

Light Conv [16] is used to unify the number of channels after feature extraction, making it easier for subsequent feature fusion.

### 2.2.1. Channel Attention

The attention mechanism can effectively improve the sensitivity of the module to the fault target and make it pay attention to more effective features. Due to the strong background interference of the fault image, we introduce the Channel Attention module in C2f. The Channel Attention sub-module is shown in Figure 3. Firstly, the feature map saves the feature information of  $C \times W \times H$  as  $W \times H \times C$  through the arrangement operation and then gains a new feature map through the inverse arrangement and activation function after passing through the perceptron.



#### Figure 3. Channel Attention.

The output of each bottleneck of C2f will be concat because the low-level feature map contains more detailed information but lacks semantic and context information. At the same time, the high-level feature map has richer semantic information and context information. Then, we introduce the channel attention mechanism before the fusion of low-level feature maps and high-level feature maps, which can improve the accuracy and robustness of target detection. The improved C2f\* is shown in Figure 4.



Figure 4. Improved C2f\*.

#### 2.2.2. Triplet Attention

The feature extraction process for small target feature maps is typically intricate, necessitating the establishment of inter-dependencies between spatial and channel dimensions. This article employs Triplet Attention as a novel method to compute attention weights by capturing cross-dimensional interactions through a three-branch structure. It can improve the restriction of the input feature map or focus on the area of concern through multidimensional interaction, as well as its inherent ability to broadly understand more contextual information that is discriminative for a certain target class.

Figure 5 shows the three branches of the Triple Attention. Within the first branch, we utilize pooling technology, the convolution layer, and the activation layer with no rotation in any dimension. We employ  $7 \times 7$  filters as they have an effective receptive field and cover more areas on the input feature map to extract features. In the second branch, the interaction is carried out over the W axis. We rotate the input feature 90 degrees counterclockwise and perform the same operation as branch one. Finally, we rotate the output 90 degrees clockwise around the W axis to maintain the original shape of the feature. In the third branch, the interaction is carried out over the H axis, and the reactions between the channel and spatial dimensions have been established. Then, the rotation features are averaged and pooled to obtain two feature maps of spatial dimension. Then, the features are combined by the convolution layer and the activation function to maintain the original shape of the features.



Figure 5. Triplet Attention (TA) network structure.

The article improves the original TA by changing the convolution layer to a lighter Light Conv structure, which achieves lightweight by using the combination of deep separable convolution and point-by-point convolution. In the activation layer, ReLU is used to activate function instead of Sigmoid to reduce the amount of calculation and achieve faster convergence.

# 2.2.3. Light Conv

Light Conv in YOLOv8 is a lightweight convolution operation that can be used to reduce the number of parameters and calculations of the model. Light Conv realizes lightweight by using a combination of deep separable convolution and pointwise convolution, as shown in Figure 6. Deep separable convolution [25] is calculated through  $1 \times 1$  and then decomposes the standard convolution into multiple convolutions. The number of convolutions is equal to the number of channels. Finally, concat is performed.



Figure 6. Light Conv model.

#### 2.3. Improved Neck Network

The backbone network is utilized for feature extraction, while the neck structure is responsible for fusing the extracted feature vectors. In the case of multi-scale object detection, small target objects inherently possess fewer pixels, and with subsequent downsampling, their features are more prone to be lost. To address this issue, the BiFPN fusion mechanism is introduced to enhance the detection capability of multi-scale objects.

# 2.3.1. BiFPN Network

FPN [26] is the earliest algorithm to propose feature fusion in the direction of object detection. Its structure can be divided into three parts: bottom-up pathway, top-down pathway, and lateral connections. Multiple levels of FPN have their own output layers, and each output layer has a different scale receptive field. YOLOv3 of the YOLO series

first adopts this feature pyramid structure, which provides a more powerful semantic information capture capability for the feature map of each layer by fusing the deep semantic information and the shallow texture information. However, this top-down FPN network is limited by one-way information flow. In recent years, numerous approaches have been proposed to enhance FPN. The neck structure of YOLOv4 and YOLOv5 incorporates the Path Aggregation Network (PANet) [27] structure, which introduces a bottom-up pathway in addition to FPN. Three models are shown in Figure 7. YOLOv8 also leverages the PANet structure for network feature fusion, but it removes two convolutional connection layers and adjusts the block count of the C2f module.



Figure 7. Three types of fusion structure.

The BiFPN [21] structure used in this article is an improvement on the basis of FPN and PANet. The node with only one input edge and no feature fusion is deleted, which simplifies the two-way network without affecting the feature network. If the original input node and output node are on the same layer, we will add an additional edge between the two. That is—jump connection, which can integrate more features without adding too much cost. In order to fuse features with different resolutions, BiFPN adds an additional weight. The fast normalized fusion Formula (1) is adopted to keep the weight at 0–1, which improves the running speed of the model on GPU, and the formula is as follows:

$$O = \sum_{i} \frac{\omega_{i}}{\varepsilon + \sum_{j} \omega_{j}} \cdot I_{i}$$
(1)

 $\omega_i$ —weight size that can be learned;

 $I_i$ —input features size;

 $\epsilon$ —a small value to avoid numerical instability.

Based on the idea of BiFPN, this article extracts the features of the YOLOv8s network with a size  $40 \times 40$ ,  $80 \times 80$  feature map being connected across layers, which can more accurately locate and obtain feature information and increase the ability of the network to detect fault targets in complex backgrounds. The specific network structure is shown in Figure 8.



Figure 8. Bidirectional Feature Pyramid Network (BiFPN) feature fusion network.

# 2.3.2. Improved BiFPN Fusion Mode

In addition to the fast normalized fusion, the fusion structure in the BiFPN network can also carry out expanded convolution through different expanded convolution rates. When the convolution kernel size is  $3 \times 3$ , and the expanded convolution rate is 1, 3, and 5, there can be fusion modes (a), (b), and (c), as shown in Figure 9 below. Methods (a) and (c) are weighted fusion and concatenation operations, respectively. They add feature mapping on the spatial dimension and channel dimension directly. Method (b) is a self-adaptive fusion method. Specifically, assuming that the size of the input can be expressed as (bs, C, H, W), we can obtain the spatial adaptive weight of (bs, 3, H, W) through convolution concatenation and SoftMax. The three channels correspond to three inputs in turn, and their outputs can aggregate context information by calculating weighting. It can be concluded from the literature [21] that (c) is more suitable for small object detection, while method (b) has the greatest improvement for large and medium-sized targets, and the improvement brought by method (a) is basically between the two. In the experimental part of this article, the three fusion methods will be compared with BiFPN, and the fusion method that most conforms to the data set in this article will be obtained.



Figure 9. Three types of fusion methods.

# 2.4. WIoU

The tag allocation of YOLOv8s adopts a dynamic matching strategy, which is used to allocate tags to the anchor of the ground truth feature map constructed by calculating loss. The loss part is changed from anchor-based to anchor-free, and the loss function of DFL [22] and Complete-IoU (CIoU) [28] is introduced as the regression branch, which makes the classification and regression tasks have high consistency.

Transmission line fault detection belongs to multi-category detection tasks, and the difficulty of classification and position in the detection phase is greater than that of single-category detection, so the positioning ability of the network should be strengthened. Therefore, this article uses the combination of distributed focus loss function DFL improved by focus loss and WIoU [23] as regression loss.

DFL is to optimize the probabilities of the two positions closest to the label y in the form of cross-entropy so that the network can focus on the distribution of the adjacent area of the target position faster. These two positions can be defined as  $y_i$  and  $y_{i+1}(y_i < y < y_{i+1})$ . Therefore, DFL is shown in Formula (2).

$$DFL(S_i, S_{i+1}) = -((y_{i+1} - y)\log(S_i) + (y - y_i)\log(S_{i+1}))$$
(2)

 $s_i(s_{i+1})$ —ensure that the estimated regression target *y* is infinitely close to label *y*.

IoU represents the overlap ratio between the predicted box and the real box, which refers to the ratio of the intersection and union of the two boxes. We will record the corresponding position of the anchor box (x, y, w, h) in the target box as  $(x_{gt}, y_{gt}, w_{gt}, h_{gt})$ . Assuming that the predicted box overlaps with the real box, then Figure 10 shows the minimum closed box and the connection between the center point.



Figure 10. Closed box and center point connection.

WIoU is an optimized strategy for bounding box loss and classification loss. When the training dataset contains low-quality examples, geometric factors such as distance and aspect ratio can increase the penalty on low-quality examples, which then affects the generalization of the model. WIoU v1 builds a distance attention mechanism to prevent slow convergence through a two-layer attention mechanism, reducing the harmful gradient generated by low-quality examples thereby improving the generalization of the model. $R_{WIoU}$  represents the loss of a high-quality anchor frame, the WIoU v1 formula is as follows:

$$L_{WIoUv1} = R_{WIoU}L_{IoU} \tag{3}$$

$$R_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right)$$
(4)

 $L_{IoU}$ —the degree of overlap between the predicted box and the actual box;

 $W_g$ —width of the minimum enclosed box;

 $H_g$ —height of minimum enclosed box;

 $x_{gt}$ —the abscissa of the center point of the real box;

 $y_{gt}$ —the vertical coordinate of the center point of the real box.

On this basis, to avoid large harmful gradients generated by low-quality samples, a small gradient gain is introduced to focus the bounding box regression on the anchor box of normal quality. WIoU v2 and WIoU v3 have incorporated monotonic and non-monotonic focus mechanisms based on the construction of bounding box losses, respectively. WIoU v2 constructs monotonic focusing coefficients  $L_{IoUI}^{\gamma*}$ ; the formula is as follows:

$$L_{WIOUv2} = L_{IOU}^{\gamma*} L_{WIOUv1}, \gamma > 0$$
<sup>(5)</sup>

 $L_{IoU}^{\gamma^*}$ —the gradient gain will decrease as the overlap decreases.

WIOU v3 constructs a non-monotonic focusing mechanism by defining outliers  $\beta$  to describe the quality of the anchor frame and a non-monotonic focusing coefficient is constructed by using outliers and being applied to WIoU v1, as shown below:

$$L_{WIOUv3} = rL_{WIOUv1} \tag{6}$$

$$r = \frac{\beta}{\delta \alpha^{\beta - \delta}} \tag{7}$$

*r*—non-monotonic focusing coefficient;

 $\alpha$ ,  $\gamma$ —hyper-parameters.

WIoU uses a dynamic non-monotonic focusing mechanism to evaluate the quality of anchor frames and uses gradient gain to construct attention-based bounding box loss. This article will conduct comparative experiments on the three types of loss functions v1, v2, and v3 in the experimental analysis section to obtain the most suitable loss function for the dataset in this article.

#### 2.5. Summary

This chapter focuses on the improvement of the benchmark model YOLOv8. Firstly, the use of improved Mosaic data augmentation increases the sample of each batch of image training, resulting in higher training efficiency. Secondly, in the feature extraction part, the improved TA structure is added to calculate attention weights, which are applied to key features. Next, two skip connections are added in the feature fusion part to perform cross-layer fusion on feature maps of sizes 80 and 40, and new fusion methods and deformable convolutions are introduced to compare and observe network performance. Finally, the WIoU optimization is introduced to optimize the bounding box loss and classification loss, improving the generalization performance of the model.

#### 3. Experimental Testing and Analysis

#### 3.1. Experimental Environment

In this experiment, the Pytorch framework is used to call GPU for the experiment. The experimental environment and basic training parameters are shown in Table 1.

Table 1. Experimental environment.

Name	Configuration
Operating system	Windows 10
Development environment	CUDA 11.7
GPU	NVIDIA GeForce RTX 3080 Ti
Epochs	100
Batch-size	16

#### 3.2. Dataset

The dataset for transmission line fault detection is composed of images taken by unmanned aerial vehicles for inspection and data-enhanced images using image augmentation techniques (https://github.com/ppuff-lily/exxx1.git, 31 October 2023). We use the LabelImg annotation tool in the Python environment to manually annotate the collected fault images. The experiment requires identifying three types of defects: broken insulators, flashover insulators, and birds' nests, so we set up four types of labels, namely: insulator, broken, pollution-flashover, and nest.

Performing a series of random transformations on the original images through cropping, flipping, and translating can balance the dataset category. As for different label data, since both flashover and broken defects are based on insulator defects, the broken and pollution-flashover labels are then included under the anchor box of the insulator label. Compared to the nest and insulator labels, flashover and broken defects of the former (the broken and pollution-flashover labels) are smaller and relatively ambiguous and complex, and there are 1–3 flashover defects or 1–2 broken defects in one image. In order to avoid the occurrence of sample imbalance, we balance the number of labels with insulator flashover, insulator damage, and bird nest defects.

The preprocessed transmission line fault detection dataset consists of 2528 pieces of images. Firstly, the test set is divided based on a 9:1 ratio, and then the training set and validation set are divided into the same ratio from the divided 2275 images. The number of transmission line fault datasets and the number of four types of labels are shown in Tables 2 and 3.

Table 2. Dataset distribution.

Category	Quantity
Training Datasets	2047
Validation Datasets	228
Test Datasets	253

Table 3. Label distribution.

Category	Quantity/Piece
broken	1093
pollution-flashover	1592
nest	1272

# 3.3. Experimental Evaluation Index

The main evaluation indexes of the target detection algorithm include detection accuracy and model complexity. In order to comprehensively and objectively evaluate the performance of the improved YOLOv8s model, the indexes of Precision (*P*), Recall (*R*), *F*1 (*F*1-score), average Precision (*AP*), and mean average Precision (*mAP*) are measured. The specific calculation formula is as follows:

$$P = \frac{TP}{TP + FP} \tag{8}$$

$$R = \frac{TP}{TP + FN} \tag{9}$$

$$F1 = \frac{2 \cdot P \cdot R}{P + R} \tag{10}$$

$$AP = \int_0^1 P dR \tag{11}$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{12}$$

*TP*—Number of positive samples correctly identified as positive samples;

*FP*—Number of negative samples incorrectly identified as positive samples;

*FN*—Number of positive samples incorrectly identified as negative samples;

*N*—Total number of categories of detection targets;

*AP*—Area under *P*-*R* curve;

*mAP*—Average value of total *AP* of various faults detected.

# 3.4. Contrast Experiments

In order to obtain the best fusion architecture for the BiFPN network for fault detection in this article, we compare the accuracy impact of four fusion methods, "bifpn," "concat," "weight," and "adaptive," on the overall model. The experimental results are compared through five metrics: Inference, Precision, Recall, GFLOPs, and mAP50. The specific values are shown in Table 4 below.

Table	4. ]	Perf	formance	compariso	on of fo	our fusion	methods.
-------	------	------	----------	-----------	----------	------------	----------

Methods	Precision %	Recall %	mAP50 %	GFLOPs	Inference/ms
bifpn	93.2	86.6	91.5	52.2	1.9
concat	90.0	90.4	91.6	56	2.3
weight	89.2	89.7	91.1	57.6	2.5
adaptive	90.8	90.3	91.0	60.2	3.3

According to the analysis based on the data below, the highest Precision among the four types of fusion structures is the "concat" structure, with an mAP of 91.6%; then the "bifpn" structure is followed, which reaches 91.5%, although mAP in the "bifpn" structure is slightly lower than the previous structure, its Inference and GFLOPs are the lowest. After fusing  $80 \times 80$  and  $40 \times 40$  feature maps across layers, the depth and width of the network

model will gradually increase. Therefore, this article chooses the relatively lightweight "bifpn" as the fusion method for the neck network.

Figure 11 shows the heatmap of the Triplet Attention output after passing through P9. The color regions indicate the feature intervals that the network focuses on. The redder the color, the higher the degree of attention of the network. At this point, the network gradually focuses its attention on defect features.



Figure 11. Comparison of thermal diagrams.

Since the classification and positioning of multi-category detection are more difficult than those of single-category detection, the positioning ability of the network should be strengthened. The WIoU used in this article has three versions: v1, v2, and v3. WIoU v1 constructs a boundary box loss based on attention, while WIoU v2 and v3 add a focus mechanism by constructing a gradient gain calculation method. v2 adopts a monotonic focus mechanism, while v3 adopts a non-monotonic focus mechanism. Through comparison of four metrics, Precision, Recall, Inference, and mAP, the specific values are shown in Table 5 below. Due to the addition of BiFPN, our model has a large network depth, so we use WIoU v2 with a monotonic focusing mechanism to optimize the bounding box loss and classification loss.

Table 5. Comparison of three types of losses.

Methods	Precision %	Recall %	mAP50 %	Inference/ms
WIoU-v1	90.8	89.5	92	1.9
WIoU-v2	91.6	88.8	92.1	1.8
WIoU-v3	92.3	86.6	91.5	1.8

# 3.5. Ablation Experiments

All ablation experiments are conducted on the same dataset, and all convolutional training starts from scratch without using weight files. To verify the impact of the proposed improved module on detection performance, we conduct ablation experiments using YOLOv8s as a benchmark model, including TA, BiFPN, and WIoU.

Detection results of the ablation study detection model are shown in Table 6. The improved model mAP increases from 88.3% to 92.4%, and Precision has also been improved from 89.8% to 92.1%. The addition of Triplet Attention calculates attention weights and applies them to key features. Additionally, the two cross-layer connections are introduced in the improved BiFPN, specifically target feature maps with sizes of  $80 \times 80$  and  $40 \times 40$ , resulting in improved detection accuracy for the three defects. Lastly, the incorporation of the WIoU loss function optimizes bounding box loss and classification loss and improves model performance.

Group	YOLOv8	TA	BiFPN	WIoU	Precision %	Recall %	mAP50 %	Inference/ms
1	$\checkmark$				89.8	85.9	88.3	1.8
2					89.2	88.5	90.7	1.7
3					90.3	91	91.6	2.2
4	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	92.1	88.4	92.4	2.3

Table 6. The detection results of ablation study detection models in this study.

In addition to verifying the effectiveness of the three improved algorithms through fusion experiments, this article visualizes the regression loss data during the training process, as shown in Figure 12. We use different symbols to position the data every ten epochs. It can be clearly seen that the introduction of WIoU greatly optimizes the bounding box regression loss.



Figure 12. Comparison of training loss curves.

3.6. Verification of Prediction Results

#### 3.6.1. Model Performance Analysis

To clearly reflect the detection effect of the improved algorithm in power transmission line fault detection in this article, we plot the training loss and validation loss after training in the same graph to analyze the performance of the model. Figures 13 and 14 show the classification loss and DFL loss of the original model and the improved model during the training and validation rounds, respectively. From the graphs, it can be seen that the coincidence degree of the improved model training and verification curve is better than that of the original model, indicating that the improved model can better prevent the occurrence of overfitting and has good performance.

We use 5-fold cross-validation to evaluate the robustness of the model and the quality of the data. We randomly divided the 2528 images into five groups: 505, 507, 505, 506, and 505. We convolve the five sets of data into the improved model, and the results are shown in Table 7. Although the accuracy of the model trained on the reduced dataset is not as high as that on the complete dataset, the mAP of the five sets of data is not significantly different, indicating that the improved model has good robustness. However, the effect on the reduced dataset is not as good as that on the complete dataset, which may be due to the low quality of individual images. In the future, we will focus on solving possible problems to improve the model performance.



Figure 13. YOLOv8 CLS and DFL comparison.



Figure 14. Improved YOLOv8 CLS and DFL comparison.

Table 7. Five-fold cross-validation
-------------------------------------

Groups	Precision %	Recall %	mAP50 %	F1 %
1	90.9	83.8	88.2	87.2
2	90.2	86.9	89.9	88.5
3	90.3	84.8	90.0	87.5
4	90.8	83.9	89.1	87.2
5	89.7	84.6	89.0	87.1

3.6.2. Defect Detection Results

Table 8 shows the detection results of 253 defect images in the test set. The mAP of all defects of the improved model is higher than that of the original model, in which the mAP of pollution-flashover defect increases from 67.3% to 79%, and the mAP of broken and nest are also improved. Therefore, the total map rises from 84.2% to 90.1%, with an increase of 5.9%. It shows that the improved YOLOv8 model can identify and locate defects more effectively.

Table 8. Detection results of various defects.

	Improved YOLOv8s			YOLOv8s		
Defect	Precision %	Recall %	mAP %	Precision <sup>(Val)</sup> %	Recall %	mAP %
Total	86.7	88.4	90.1	86.2	83.5	84.2
broken	81.7	88.7	89.5	80.7	79.2	81.6
insulator	97.7	100	99.5	98.2	99.8	99.5
pollution-flashover	86.1	72.9	79	84.7	63	67.3
nest	81.4	91.9	92.3	81.1	92.1	88.6

Since dealing with multiple categories of failures, class-specific metrics are used to verify the performance of the model on each individual class. Table 9 shows the Precision, Recall, mAP, and Inference of the three defects in the improved model. The detection accuracy of the single category is higher than that of multiple categories in Table 8, in which the mAP of broken is improved by 1.9%, and the mAP of nest and pollution-flashover are improved by 7.2% and 8.6%; this shows that our model still has good performance for single-category detection.

Table 9. Class-specific metrics.

Class	Precision %	Recall %	mAP50 %	Inference/ms
broken	94.3	86.5	91.4	3.3
pollution-flashover	89.5	76.6	87.6	3.7
nest	99.9	100	99.5	3.0

# 3.6.3. Visualization of Defect Characteristics

To verify the detection performance of the proposed algorithm, we randomly select pollution-flashover, broken, and bird nest defects in the dataset for comparison experiments in a complex background. The detection results are shown in Figure 15. In the case of dense defect targets in a complex background, the original YOLOv8s model has the problem of low accuracy. Under the improved YOLOv8s model in this article, the number of prediction boxes for pollution flashover defect detection ranges from 7 to 10. The detection accuracy of nest defects has increased from 87% to 91%. The detection accuracy of small target defects in complex backgrounds is also a major factor affecting interference detection accuracy. TA and BiFPN used in this article have significantly optimized the detection of small targets, with the accuracy of damage defects increasing from 41% to 69% in Figure 15 below.



Figure 15. Detection under different defects.

In order to observe the location information of defects more conveniently, we use a heatmap to visualize feature extraction, as shown in Figure 16. The color on the improved YOLOv8s is redder on the heatmap, indicating that the improved algorithm has enhanced attention to defects. Therefore, the improved algorithm can be well applied to transmission line fault detection tasks under complex backgrounds.



Figure 16. Comparison of thermodynamic diagrams under different defects.

### 4. Discussion and Conclusions

This article analyzes the impact of complex backgrounds in images collected by electric power inspection on the training of YOLOv8, showing that the extraction of multi-scale discriminative features and cross-layer fusion can improve detection accuracy. At the same time, in order to make power inspection more comprehensive in detecting faults, we have added birds' nests for multi-scale fault detection.

In order to improve the accuracy of defect detection for transmission line insulators and birds' nests, this article mainly makes improvements in three aspects: (1) we propose an improved TA structure to calculate new attention weights, which enhances the network's ability to extract discriminative features. (2) Using two cross-layer connections to fuse more information, thereby improving the model's accuracy in detecting defects in complex backgrounds. (3) Since the tasks performed belong to multiple categories, we introduce WIoU v2 to optimize the bounding box loss and classification loss, allowing the model to focus on difficult examples and obtain better model performance. The experimental results show that the mAP of the improved YOLOv8 in transmission line fault detection is 92.4%, and the model has certain advantages for insulator and bird nest fault detection under complex backgrounds.

Defect detection in transmission lines is still an emerging research area with a large number of challenges. The existing fault dataset still suffers from interference from complex backgrounds. Therefore, expanding the dataset by synthesizing data or introducing various background interferences to make the model better adapt to different complex background situations is one goal of future work. Meanwhile, further optimizing the model structure and improving the overall performance of the model is another goal of our future work.

**Author Contributions:** Conceptualization, H.Q.; methodology, H.Q. and Z.T.; software, Z.T.; validation, H.Q. and Z.T.; formal analysis, Z.T. and H.Q.; investigation, Z.T. and B.Y.; resources, H.Q.; data curation, Z.T., R.Y. and B.Y.; writing—original draft preparation, Z.T.; writing—review and editing, Z.T. and H.Q.; visualization, H.Q.; supervision, H.Q.; project administration, H.Q. and W.X.; funding acquisition, W.X. and R.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Postgraduate Research & Practice Innovation Program of Jiangsu Province under the grant number [SJCX23\_1488].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

# Abbreviations

UAV	unmanned aerial vehicle
WIoU	Wise-IoU
FPN	Feature Pyramid Network
BiFPN	Bidirectional Feature Pyramid Network
DFL	Distribution Focal Loss
CIoU	Complete-IoU

#### References

- 1. Chaitanya, B.K.; Yadav, A. An intelligent fault detection and classification scheme for distribution lines integrated with distributed generators. *Comput. Electr. Eng.* **2018**, *69*, 28–40. [CrossRef]
- Nyangaresi, V.; Jasim, H.; Mutlaq, K.; Abduljabbar, Z.; Ma, J.; Abduljaleel, I.; Honi, D. A Symmetric Key and Elliptic Curve Cryptography-Based Protocol for Message Encryption in Unmanned Aerial Vehicles. *Electronics* 2023, 12, 3688. [CrossRef]
- 3. Kezunovic, M. Smart fault location for smart grids. IEEE Trans. Smart Grid. 2011, 2, 11–22. [CrossRef]
- Odo, A.; Mckenna, S.; Flynn, D.; Vorstius, J.B. Aerial image analysis using deep learning for electrical overhead line network asset management. *IEEE Access* 2021, 9, 146281–146295. [CrossRef]
- 5. Anjum, S.; Jayaram, S.; El-Hag, A.; Jahromi, A. Detection and classification of defects in ceramic insulators using RF antenna. *Trans. Dielectr. Electr. Insul.* **2017**, *24*, 183–190. [CrossRef]
- 6. Liu, X.; Miao, X.; Jiang, H.; Chen, J. Data analysis in visual power line inspection: An in-depth review of deep learning for component detection and fault diagnosis. *Annu. Rev. Control* 2020, *50*, 253–277. [CrossRef]
- Biagetti, G.; Crippa, P.; Falaschetti, L.; Turchetti, C. A Machine Learning Approach to the Identification of Dynamical Nonlinear Systems. In Proceedings of the European Signal Processing Conference, A Coruna, Spain, 2–6 September 2019; pp. 1–5.
- Xu, B.; Zhao, Y.; Wang, T. Development of power transmission line detection technology based on unmanned aerial vehicle image vision. SN Appl. Sci. 2023, 5, 1–15. [CrossRef]
- Ma, Y.; Li, Q.; Chu, L.; Chu, L.; Zhou, Y. Real-time detection and spatial localization of insulators for UAV inspection based on binocular stereo vision. *Remote Sens.* 2021, 13, 230. [CrossRef]
- 10. Chen, K.; Huang, C.; He, J. Fault detection, classification and location for transmission lines and distribution systems: A review on the methods. *High Volt.* **2016**, *1*, 25–33. [CrossRef]
- 11. He, N.; Wang, S.; Liu, F.; Zhang, H.; Wu, L.; Zhou, X. Research on infrared image missing insulator detection method based on deep learning. *Power Syst. Prot. Control* **2021**, *49*, 132–140.
- 12. Lei, X.; Sui, Z. Intelligent fault detection of high voltage line based on the Faster R-CNN. *Measurement* **2019**, *138*, 379–385. [CrossRef]
- 13. Liu, C.; Wu, Y.; Liu, J.; Sun, Z. Improved YOLOv3 Network for Insulator Detection in Aerial Images with Diverse Background Interference. *Electronics* **2021**, *10*, 771. [CrossRef]
- 14. Wu, T.; Wang, W.; Yu, L.; Xie, B.; Yin, W.; Wang, H. Insulator Defect Detection Method for Lightweight YOLOV3. *Comput. Eng.* **2019**, 45, 275–280.
- 15. Wu, Z.; Wu, Z.; Sun, S. Multiple Defect Detection of Insulators Based on Improved YOLOv5 Algorithm. *High Volt. App.* **2023**, 1–11. Available online: https://link.cnki.net/urlid/61.1127.TM.20230814.1815.002 (accessed on 1 November 2023).
- 16. Jacob, S.; Francesco. What Is YOLOv8? the Ultimate Guide. Available online: https://blog.roboflow.com/whats-new-in-yolov8/ (accessed on 1 November 2023).
- 17. Chen, Y.; Zhang, S.; Ran, X.; Wang, J. Aircraft Target Detection Algorithm Based on Improved YOLOv8 in SAR Image. *Telecommun. Eng.* **2023**, *84*, 1–8.
- Gao, A.; Liang, X.; Xia, C.; Zhang, C. A dense pedestrian detection algorithm with improved YOLOv8. J. Graph. 2023, 1–9. Available online: https://kns.cnki.net/kcms2/detail/10.1034.T.20230731.0913.002.html (accessed on 1 November 2023).
- Bochkovskiy, A.; Wang, C.; Liao, H. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* 2020, arXiv:2004.10934v1.
   Shimaa, S.; Khalid, A.; Paweł, P.; Ryszard, T.; Mohamed, H. Graph convolutional network with triplet attention learning for person re-identification. *Inf. Sci.* 2022, *617*, 331–345.
- 21. Tan, M.; Pang, R.; Le, Q. EfficientDet: Scalable and Efficient Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10778–10787.
- 22. Li, X.; Wang, W.; Wu, L.; Chen, S.; Hu, X.; Li, J.; Tang, J.; Yang, J. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 21002–21012.
- 23. Tong, Z.; Chen, Y.; Xu, Z.; Yu, R. Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism. *arXiv* 2023, arXiv:2301.10051.
- 24. Dai, Z. Uncertainty-aware accurate insulator fault detection based on an improved YOLOX model. *Energy Rep.* 2022, *8*, 12809–12821. [CrossRef]
- 25. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. arXiv 2017, arXiv:1610.02357v3.
- Lin, T.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 936–944.

- 27. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
- 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, 3 April 2020; pp. 12993–13000.

**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.