



Article

A Defect Detection Method Based on BC-YOLO for Transmission Line Components in UAV Remote Sensing Images

Wenxia Bao ¹, Xiang Du ¹, Nian Wang ^{1,*}, Mu Yuan ¹ and Xianjun Yang ²¹ School of Electronics and Information Engineering, Anhui University, Hefei 230601, China² Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, China

* Correspondence: wn_xlb@ahu.edu.cn

Abstract: Vibration dampers and insulators are important components of transmission lines, and it is therefore important for the normal operation of transmission lines to detect defects in these components in a timely manner. In this paper, we provide an automatic detection method for component defects through patrolling inspection by an unmanned aerial vehicle (UAV). We constructed a dataset of vibration dampers and insulators (DVDI) on transmission lines in images obtained by the UAV. It is difficult to detect defects in vibration dampers and insulators from UAV images, as these components and their defective parts are very small parts of the images, and the components vary greatly in terms of their shape and color and are easily confused with the background. In view of this, we use the end-to-end coordinate attention and bidirectional feature pyramid network “you only look once” (BC-YOLO) to detect component defects. To make the network focus on the features of vibration dampers and insulators rather than the complex backgrounds, we added the coordinate attention (CA) module to YOLOv5. CA encodes each channel separately along the vertical and horizontal directions, which allows the attention module to simultaneously capture remote spatial interactions with precise location information and helps the network locate targets of interest more accurately. In the multiscale feature fusion stage, different input features have different resolutions, and their contributions to the fused output features are usually unequal. However, PANet treats each input feature equally and simply sums them up without distinction. In this paper, we replace the original PANet feature fusion framework in YOLOv5 with a bidirectional feature pyramid network (BiFPN). BiFPN introduces learnable weights to learn the importance of different features, which can make the network focus more on the feature mapping that contributes more to the output features. To verify the effectiveness of our method, we conducted a test in DVDI, and its mAP@0.5 reached 89.1%, a value 2.7% higher than for YOLOv5.

Keywords: vibration dampers; insulators; feature fusion; YOLOv5; defect detection; attention mechanism



Citation: Bao, W.; Du, X.; Wang, N.; Yuan, M.; Yang, X. A Defect Detection Method Based on BC-YOLO for Transmission Line Components in UAV Remote Sensing Images. *Remote Sens.* **2022**, *14*, 5176. <https://doi.org/10.3390/rs14205176>

Academic Editors: Qian Du, Yanni Dong and Xiaochen Yang

Received: 2 September 2022

Accepted: 13 October 2022

Published: 16 October 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Insulators and vibration dampers are exposed to the outdoor environment for long periods of time, are subject to weather and high mechanical tension, and are therefore likely to fail physically [1,2]. The primary function of insulators is to support wires and prevent the current from returning to ground, and when insulators have problems such as fractures or cracks, they are easily pierced, resulting in zero insulation resistance at both ends of the insulator string. Insulation is then lost, leading to power interruptions and outages [3]. The primary function of vibration dampers is to protect power lines from wind-induced periodic vibrations, thus reducing accidents such as transmission line fatigue and segmental strands [4]. When vibration dampers are exposed to the environment for a long period, they tend to rust. Furthermore, rusting may cause the steel strands to loosen, and the hammer head of the vibration dampers will then easily deform, slide, and fall off [5,6], which affects power transmission. The timely detection and replacement

of damaged insulators and vibration dampers can therefore guarantee the effective and normal operation of the transmission lines.

Transmission line inspection is an important component of a power system, and the most primitive form is manual inspection. However, traditional manual inspection requires significant human and material resources, and the detection speed is slow. With the continuous development of science and technology, modern power line inspection involves helicopter [7], robot [8], and unmanned aerial vehicle (UAV) inspection methods [9] rather than the original manual methods. Of these, UAV inspection has gradually become an important part of transmission line inspection due to its low cost, high efficiency, and flexibility [10]. However, in the transmission line images taken by UAVs, the vibration dampers are relatively small, and defects such as rust are easily confused with the background color. Although the insulators are much larger than vibration dampers, the location of their self-destruction defect may only be a small piece, so it is also difficult to locate on longer insulation strings in the remote sensing images. Moreover, the insulators and vibration dampers are distributed in complex environments such as forests, houses, and fields, making inspection challenging. The above problems are common in UAV inspection, belonging to small target detection problems in computer vision. The existing methods are insufficient for the accurate detection of insulators and vibration dampers; therefore, this area needs to be further studied. The research we have conducted can be summarized in the following two points:

- We construct a dataset of images of vibration dampers and insulators using remote sensing images taken by UAVs, which we refer to as DVDI. There are three types of insulators (XWP, LXY, and FXBW) and four types of vibration dampers (FD, FDZ, FFH, and FR). Each type of vibration damper or insulator may have defects, or may be normal.
- We propose a defect detection method for insulator and vibration dampers, named BC-YOLO. We introduce the CA module into YOLOv5. This module embeds the location information into the channel attention by decomposing the channel attention. The CA module enhances the network's ability to detect insulators and vibration dampers in complex backgrounds. We use BiFPN instead of the original PANet feature fusion framework to better balance the feature information at different scales by weighting each scale. The BiFPN feature fusion framework enhances the network's ability to detect small targets such as vibration dampers.

The rest of this paper consists of the following chapters. Section 2 presents the relevant literature on target detection. Section 3 introduces our dataset and gives an overview of its pre-processing part. Section 4 introduces our proposed BC-YOLO network structure. Section 5 presents the related experiments and analysis. Section 6 analyzes the confusion matrix generated by training. Section 7 concludes the paper while suggesting future work.

2. Related Work

Tiantian et al. [11] proposed a feature fusion-based insulator detection method, which consisted of a histogram of directional gradient features after principal component analysis dimensionality reduction and local binary pattern features. Support vector machine (SVM) fusion features were used to build the training model. The sliding window method was used to search for candidate regions, and the non-maximum suppression method was applied to fuse the candidate windows. Finally, the location of insulator strings was calculated by linear fitting. In Reddy et al. [12], video monitoring technology is used to obtain insulator images, and discrete orthogonal s-transform and k-means clustering algorithms are used to determine the state of insulators. In Reddy et al. [13], the features of the insulator are extracted by discrete orthogonal transformation, after which the state of the insulator is estimated by SVM. Li et al. [14] based a tilt correction method on principal component analysis, which enabled them to obtain accurate feature extraction curves from insulator images. Five features were extracted from the feature curves, and finally SVM was used to identify insulators with these five features. In Zhao et al. [15],

insulators are located by azimuth angle detection and binary shape prior knowledge. The possible orientation angles of the insulator are initially detected, and for each possible angle, the insulator is retained according to its binary shape prior knowledge, and finally, all possible angles are traversed to locate the insulator. Wu et al. [16] proposed a new texture segmentation algorithm to segment complex insulator images into subregions with closed, smooth contours. The texture features of the insulators were extracted using a gray level co-occurrence matrix and calculated using a fast gray level co-occurrence integrated algorithm. In Liao et al. [17], the insulators are located using the local features of the insulators. Firstly, multi-scale and multi-features are introduced to represent the local features of the insulators. After that, the local features of the insulators are trained, and finally, the insulators are localized by a coarse-to-fine strategy. Jabid et al. [18] proposed a new rotation-invariant insulator detection method. Rotational invariance was achieved by an efficient method to estimate the rotation angle of all insulators in an image. Local directional map features based on sliding windows were extracted from the images, and each sliding window was classified using a support vector machine. Zhao et al. [19] proposed a method for representing the appearance of insulator string infrared images based on binary robust invariant scalable key points and vectors of local aggregation descriptors. An SVM-based classification model was integrated into a multi-scale sliding window framework for locating insulator strings in infrared images. Guifeng et al. [20] proposed a swarm optimization clustering algorithm for insulator image segmentation. Their method used an ant colony clustering algorithm for the clustering and segmentation of insulator images. Images taken during a UAV inspection are of very high resolution, but the insulators, and particularly the vibration dampers, on the transmission line are relatively small. The above methods are not effective for the detection of small target objects in high resolution images. In addition, vibration dampers may be easily confused with trees in the forest due to their color, which makes the defect detection difficult.

With the rise of deep learning techniques, more and more people are conducting target detection based on deep learning. The detection methods based on deep learning can be divided into two main categories: single-stage methods (e.g., SSD and YOLO) and two-stage methods (e.g., Faster-RCNN). In general, two-stage methods are not as fast as single-stage methods, and have larger network weight files. Li et al. [21] proposed a method for automatically detecting birds' nests on transmission lines based on Faster-RCNN. The problem of insufficient data samples was solved by zooming in on the bird nest images. Wanguo et al. [22] chose a candidate region-based SSD algorithm for defect localization and identification. The problem of insufficient samples is solved by horizontal mirroring and multi-scale training, and the network parameters are adjusted to select the appropriate network parameters for transmission line defect detection. Wu et al. [23] proposed a CenterNet-based insulator defect detection method. Blurred images are reconstructed using super-resolution in the data pre-processing part to enhance the dataset, while attention mechanisms are used in the net to reduce the interference of the background. Wu et al. [24] proposed an improved YOLOv3 algorithm for the detection of electrical connector defects. First, the K-means clustering algorithm was used to cluster the dataset and improve the detection accuracy of the defective targets. Then, single-scale feature mapping was used for detection rather than multi-scale prediction as in the original network, which not only reduced the computational effort, but also avoided false detection to some extent. Bao et al. [25] proposed a YOLOv4-based PMA-YOLO network, which adds parallel mixed attention (PMA) to the YOLOv4 network to make the network more focused on the target information. In addition, the K-means algorithm is introduced to re-cluster the anchor of vibration dampers. Finally, a multi-stage migration learning strategy is used to improve the training efficiency.

The YOLO series of target detection networks is a popular one-stage object detection method, which has made great progress in target detection by directly predicting the class and location of various objects using separate CNN networks. Compared with Regions with Revolutionary Neural Networks (RCNN) series object detection methods, YOLOv5

achieves a balance between detection speed and accuracy. In particular, YOLOv5 uses the most advanced optimization strategies in the field of CNNs, which is based on an iteration of the original YOLO series. Different degrees of optimization have been applied to data pre-processing, backbone networks, activation functions, training strategies, and anchor clustering, among others. So YOLOv5 is more suitable as a rapid benchmark network for detecting defects in insulators and vibration dampers on transmission lines. However, due to the high resolution of the remote sensing images taken by UAVs, the size of the target represented by the vibration dampers is small, and there is a large difference between the vibration dampers and insulator in size. Therefore, in this paper, a CA module is added to the structure of the backbone, and the original PANet module is replaced with Bi-FPN to improve the detection accuracy of the network. The BC-YOLO network can detect four transmission line defects (normal and damaged insulators, and normal and dislodged vibration dampers), with a mAP@0.5 reaching 89.1%.

3. Dataset

In this paper, we construct a dataset containing insulators and vibration dampers, called DVDI. All of the images in the dataset were taken by UAVs during overhead transmission line inspections by the Chinese Academy of Electric Power. The filming equipment used during the transmission line inspection was a UAV manufactured by DJI, model Phantom 4 Pro V2.0. The UAV maintains a distance of about 10 m from the transmission tower during each power inspection. The UAV model and the image during the inspection are shown in Figure 1.



Figure 1. (a) Phantom 4 Pro V2.0 UAV; (b) UAV inspection shooting data.

3.1. DVDI Dataset

There are many types of insulators and vibration dampers on transmission lines, and the corresponding type is chosen according to the usage scenario. For the most common vibration dampers, there are four types (FD, FDZ, FDY, and FFH), as shown in Figure 2. There are three common types of insulators (FXBW, LXY, and XWP), as shown in Figure 3. The DVDI dataset includes normal and defective vibration dampers as well as normal and defective insulators, as shown in Figure 4.



Figure 2. Four common types of vibration dampers: (a) FD; (b) FDZ; (c) FDY; (d) FFH.

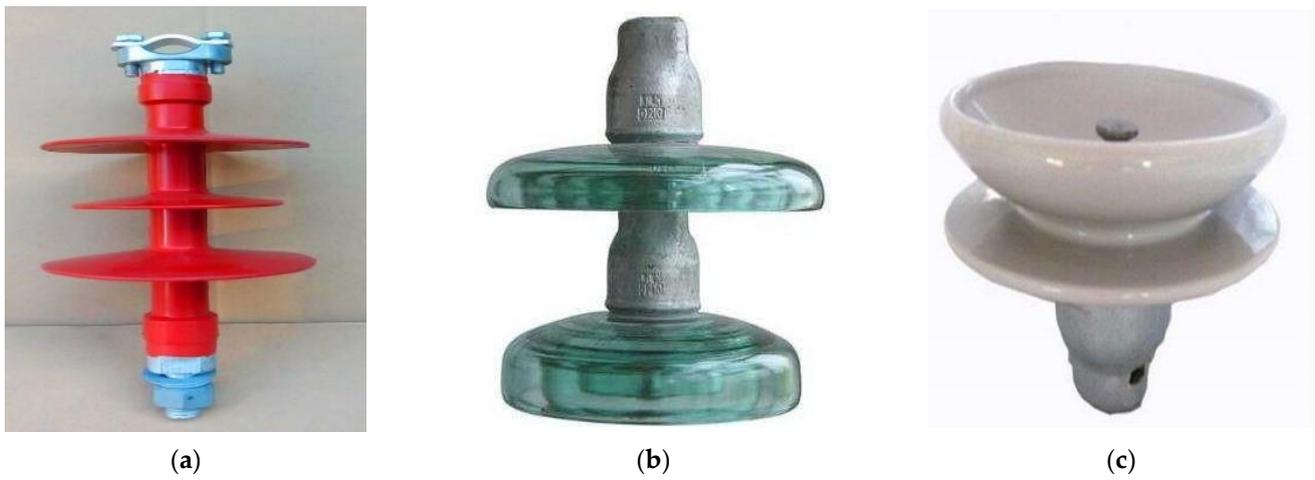


Figure 3. Three common types of insulators: (a) FXBW; (b) LXY; (c) XWP.

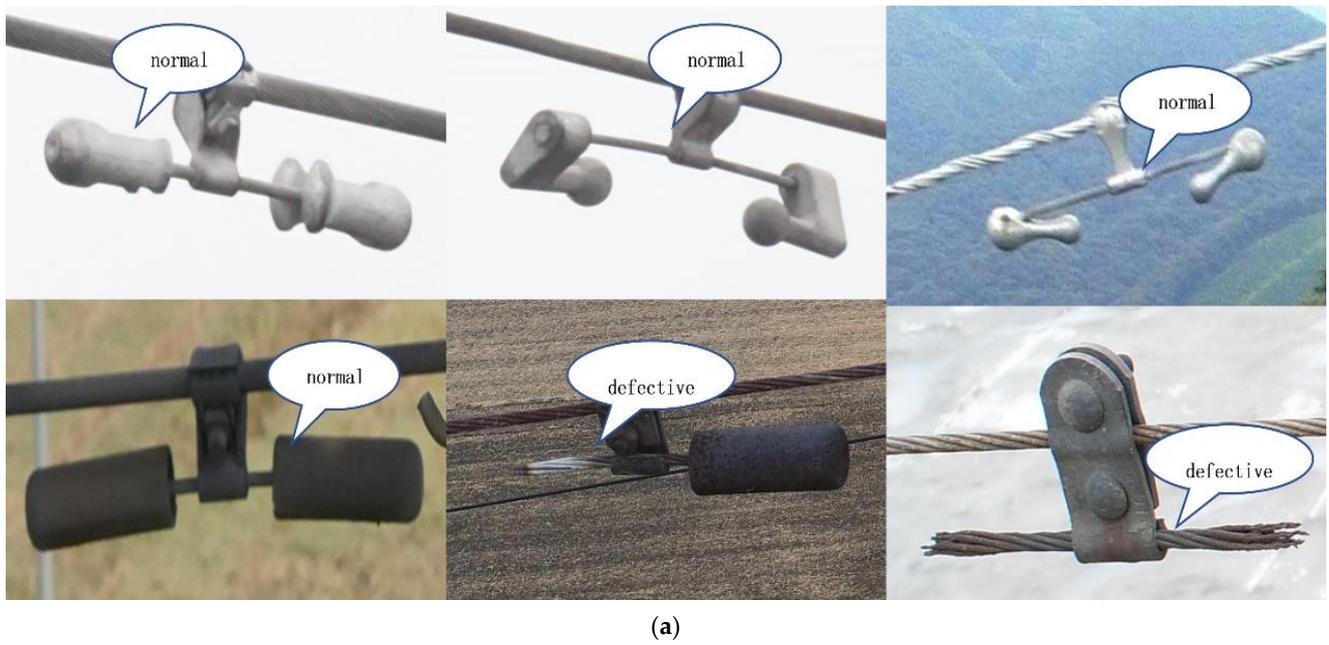


Figure 4. Cont.

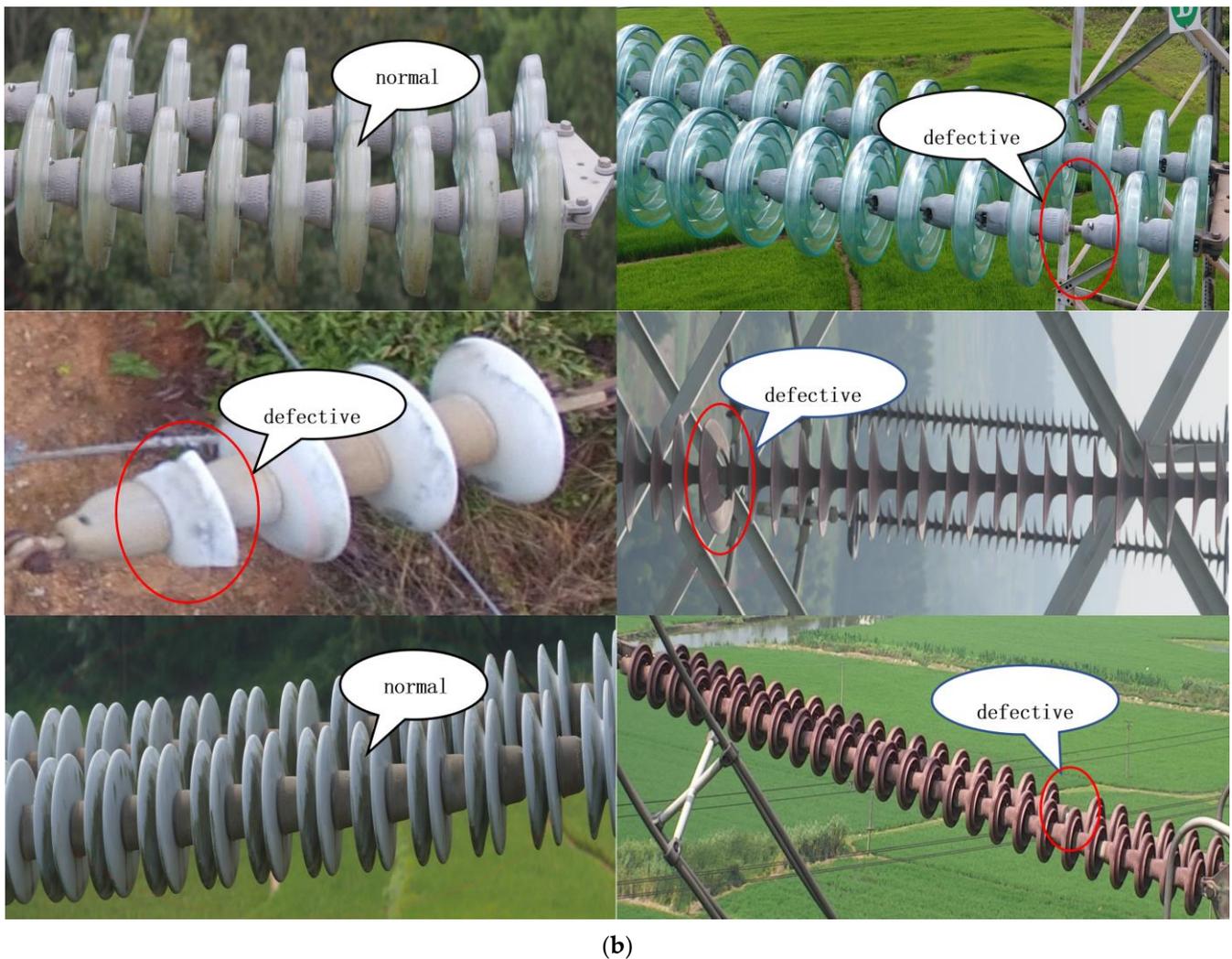


Figure 4. (a) Examples of normal and defective vibration dampers; (b) examples of normal and defective insulators.

After screening, there were a total of 976 UAV remote sensing images, which contained complex backgrounds with areas such as trees, forests, and mountains, as shown in Figure 5. It can be seen from these images that the colors of some insulators and vibration dampers mean that they are easily confused with the background, and they are therefore not easy to detect.

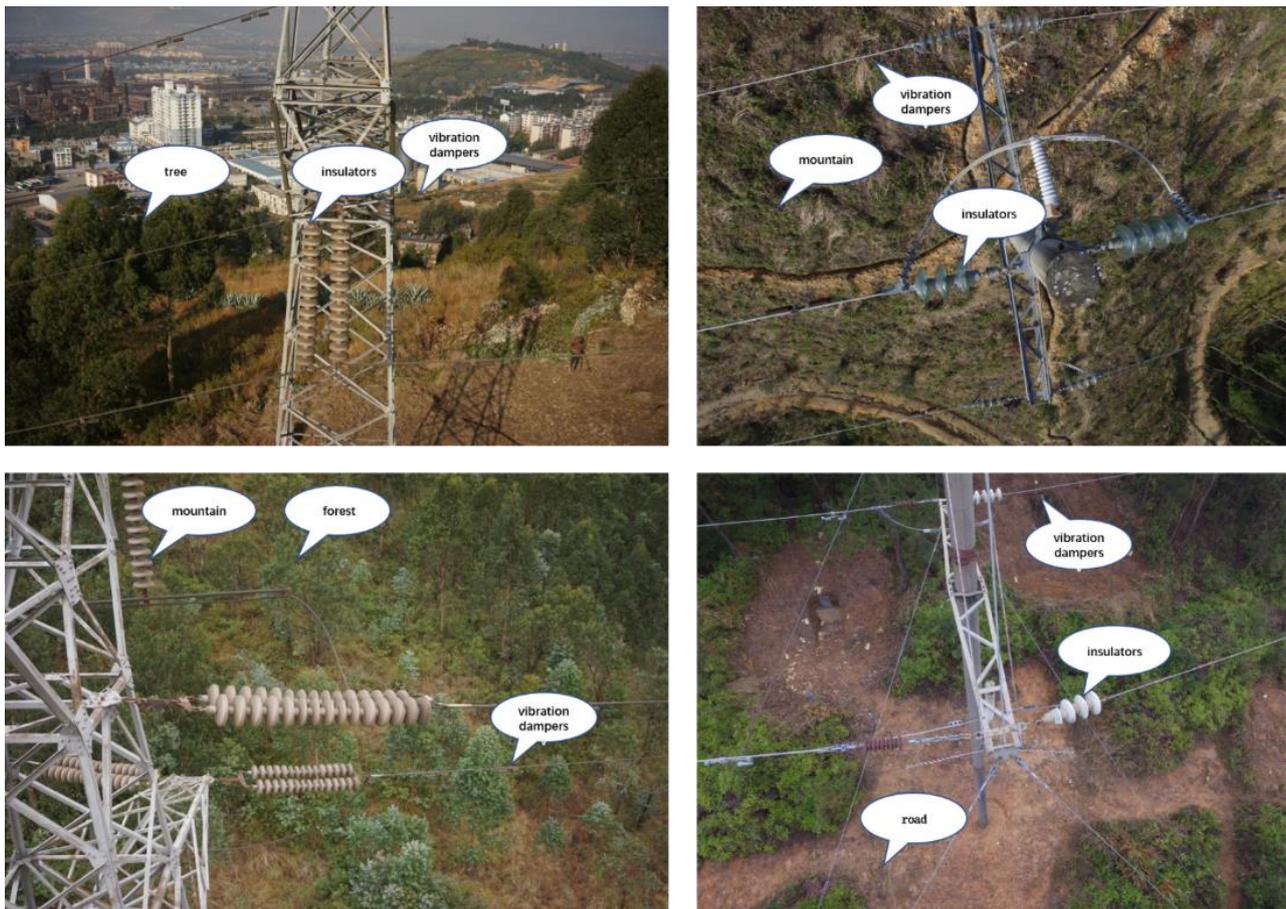


Figure 5. Partial images with vibration dampers and insulators.

3.2. Data Pre-Processing

There are 976 vibration damper and insulator images in the DVDI dataset. In order to improve the generalization ability of the models, the data augmentation techniques are often used in deep convolutional networks. These images are rotated 12° and 180° , and flipped horizontally to obtain the images of defective insulators and vibration dampers from different perspectives. The number of images in the DVDI is increased to 1500 after augmentation. The process of power inspection is affected by the intensity of the light, and insulators and vibration dampers often appear against complex mountainous backgrounds. The dual effects of light intensity and a complex environment can cause false detections and missed detections, and we therefore enhanced the contrast of the detected images to improve the overall quality. For this purpose, we used a gamma transform [26] to enhance the contrast of the image, specifically to correct images with too much or too little gray. The transformation formula is shown in Equation (1), and is a product operation that is applied to each pixel value in the original image. When $\gamma > 1$, this has a stretching effect on the histogram of the gray distribution of the image (making the grayscale stretch toward a high gray value), while a value of $\gamma < 1$ has a shrinking effect on the histogram of the gray distribution of the image (making the grayscale move toward a low gray value). Examples of pre-processed images are shown in Figures 6 and 7.

$$s = cy^\gamma \quad (1)$$

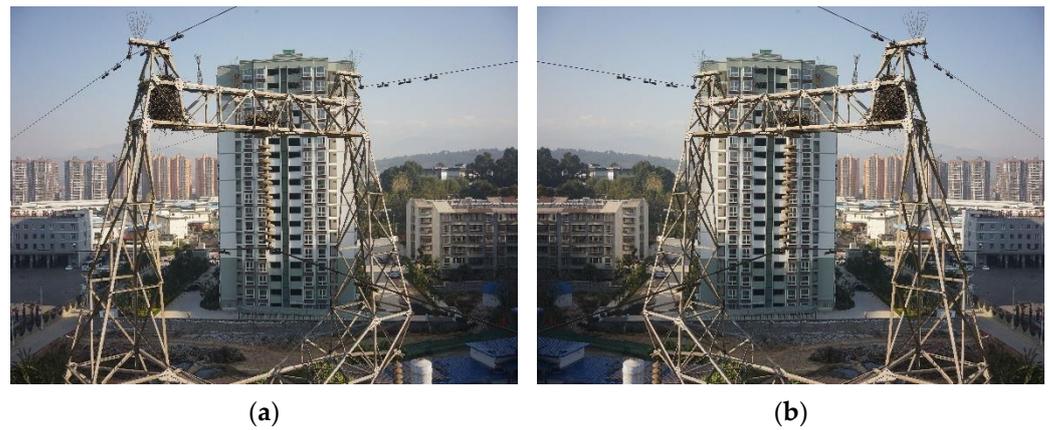


Figure 6. (a) Original image; (b) horizontal flip.



Figure 7. (a) Original image; (b) gamma transform.

3.3. Data Annotation

We used LabelImg as a labeling tool for the insulators and vibration dampers. The labeling rules were as follows: if an insulator was normal, it was labeled as 1, while if there was damage, it was labeled as 2. If a vibration damper was normal without serious bending, it was labeled as 3, while if there was shedding or serious bending, it was labeled as 4. An example of partial image annotation is shown in Figure 8. The labeled data were stored in Pascal VOC [27] format, and the labeled file format was XML. After that, we divided our dataset into a training set, validation set, and test set in the ratio of 6:2:2. The number of annotations for each class is shown in Table 1.



Figure 8. Example of annotation with LabelImg.

Table 1. Label numbers per class.

Classes	Number
1 (normal insulators)	2252
2 (damaged insulators)	632
3 (normal vibration dampers)	3610
4 (damaged vibration dampers)	636

4. Method

4.1. YOLOv5 Network Architecture

YOLOv5 is a single-stage target inspection method that locates and classifies targets by directly regressing the relative positions of candidate boxes. YOLOv5 is the latest network in a series of several iterations of the original YOLO model. Various improvements have been made in terms of data pre-processing, feature extraction, and feature fusion, have been made to greatly improve the detection accuracy of the network. A diagram of the structure of the original YOLOv5 is given in Figure 9. It has three main parts: a backbone, neck, and head. The backbone contains C3, CBL, Focus, and spatial pyramid pooling (SPP) modules [28] to extract features from the input image and pass them to the neck layer. The CBL module is composed of Convolution, Batch Normalization, and Leaky ReLU functions. The C3 module splits the input into two branches, one passing through the CBL first and then through the residual structure. After the other branch passes through the CBL, concat combines the outputs of the two branches. The SPP module is a multi-scale feature fusion stage that uses four different sizes of maximum pooling. The Focus module is a slicing operation of the image, preserving more complete information about the down-sampling of the image for subsequent feature extraction. A detailed diagram of the C3, CBL, Focus, SPP module is shown in the last panel of Figure 9. The neck layer exploits the structure of PANet [29] to produce feature pyramids. The algorithm enhances the detection capability of targets at different scales by the bidirectional fusion of low-level semantic space and high-level semantic features. The head layer consists of an anchor frame that is applied to the multi-scale feature map of the neck module that generates detection frames and divides them into corresponding categories, coordinates, and confidence levels.

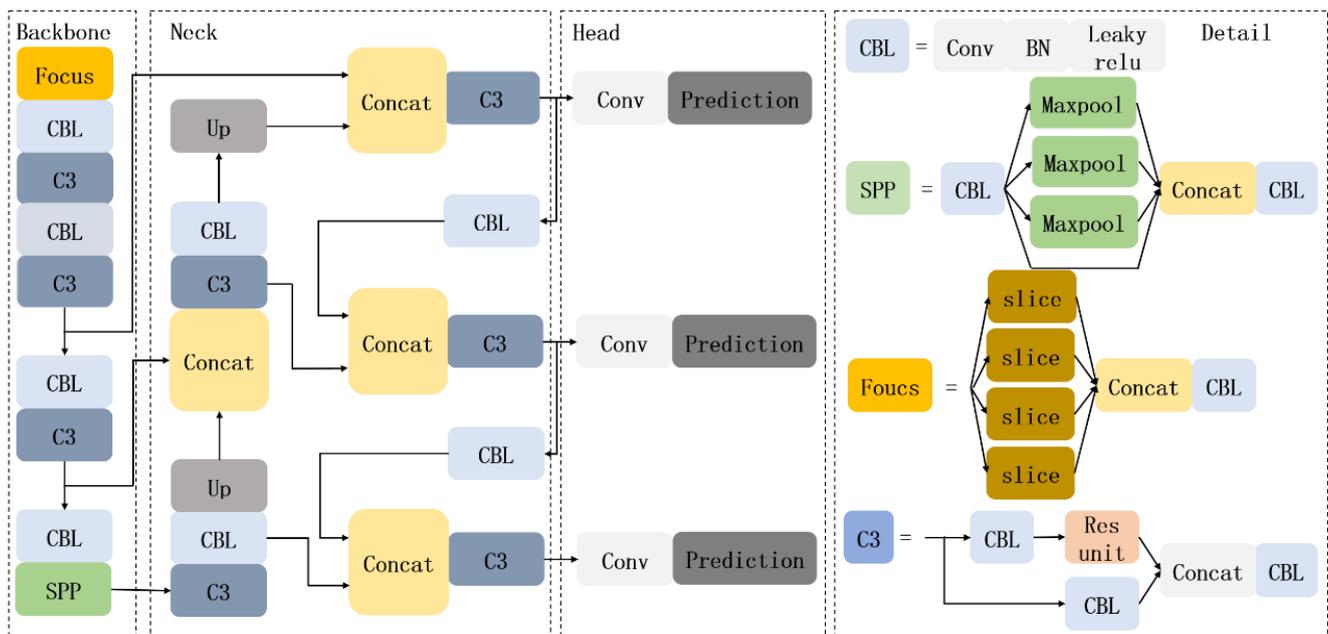


Figure 9. Architecture of the original YOLOv5 network.

4.2. Architecture of the BC-YOLO Network

The remote sensing images taken by UAVs during power inspections have complex backgrounds, and the different types of insulators and vibration dampers may have varying fault shapes and sizes. Existing methods cannot achieve accurate detection and classification of insulators and vibration dampers at the same time. Therefore, we improve on YOLOv5 and propose the BC-YOLO network. In the backbone module, we introduce a CA [30] mechanism, which can help the network to locate the detection target more accurately, thus reducing the interference of target background information. In the neck module, we introduce Bi-FPN [31] to replace the original PANet, which enhances the detection capability of the network for small targets such as vibration dampers by adding weights to each scale to adjust the contribution of each scale. A diagram of the structure of the YOLOv5 is given in Figure 10.

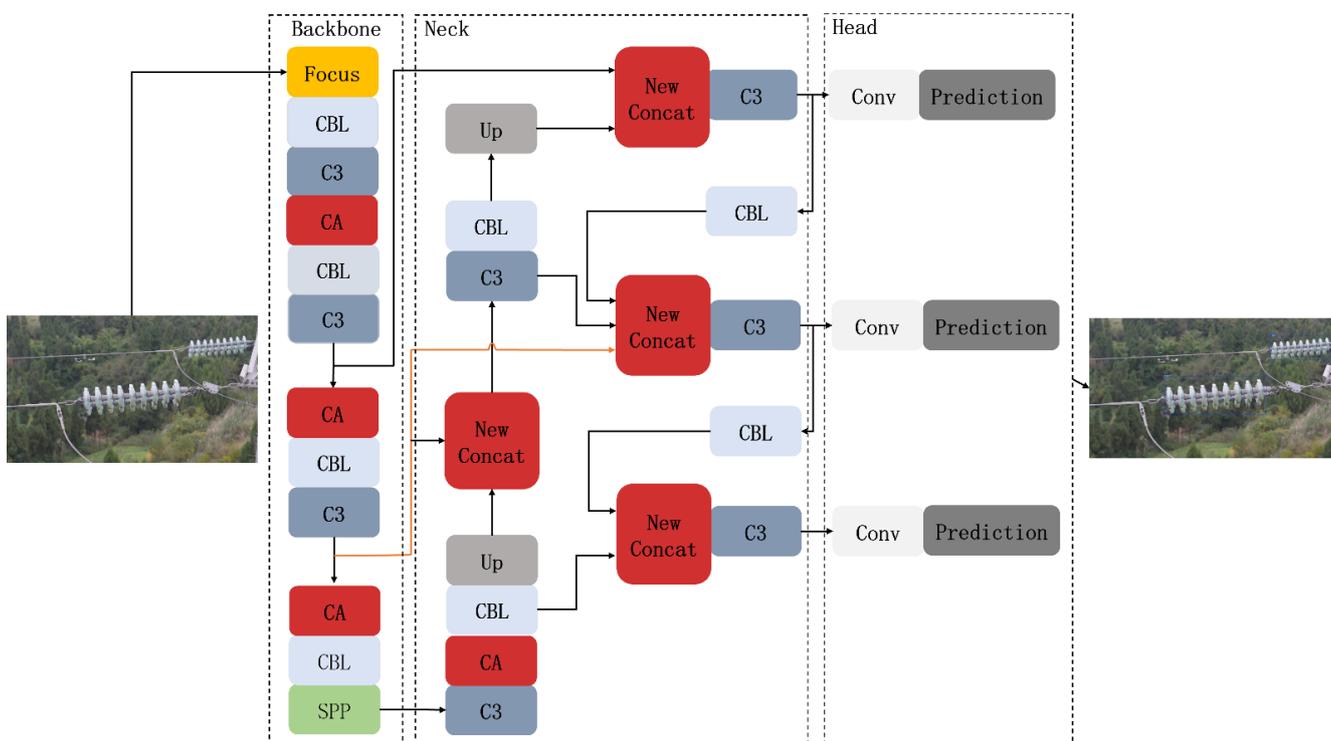


Figure 10. Architecture of the BC-YOLO network.

4.2.1. Attention Mechanism Module

The attention mechanism emerged to make the network more focused on the target information of the current task and less on other information. However, traditional attention mechanisms such as squeeze-and-excitation (SE) [32] only consider the channel information, where location information is equally important in visual targets. In view of this, the convolutional block attention module (CBAM) [33] was derived from SE, and aims to introduce location information through the use of global pooling on the channels. However, this approach can only capture local information, and cannot obtain long-range-dependent information. After several convolution layers, each position of the feature maps contains information about a local area of the original image, and CBAM is used as a weighting factor by taking the maximum and average values of multiple channels for each position, meaning that this weighting only takes into account the information on the local area. In contrast, CA attention avoids the introduction of a larger overhead by embedding the location information into the channel attention, thus allowing the mobile network to obtain information about a larger area. We therefore introduce a CA attention mechanism to improve the detection performance of our network. To solve the problem of position loss

due to global average pooling, we efficiently integrate spatial coordinate information by decomposing channel attention into two parallel 1D feature encodings. Specifically, in order to focus attention on the height and width of the image and to encode the precise location information, the input features are divided into two directions, height and width, for global averaging pooling to obtain feature maps in both the height and width directions, respectively, as shown in Equations (2) and (3). The feature maps in the height and width directions of the obtained global perceptual field are then stitched together, after which they are fed into the convolution module with a shared convolution kernel of 1×1 . Following this, the batch-normalized feature map F_1 is fed into the sigmoid activation function to obtain f , as shown in Equation (4). The feature map f is then convolved with a convolution kernel of 1×1 according to the original height and width, and the attention weights in the height and width directions are obtained after the activation function, as shown in Equations (5) and (6). Finally, the attention weights in the height and width directions and the original feature map are calculated by multiplying and weighting to obtain feature maps with attention weights in the width and height directions, as shown in Equation (7). The structure of the coordination of the attention mechanism is shown in Figure 11.

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i \leq W} x_c(h, i) \tag{2}$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j \leq H} x_c(j, w) \tag{3}$$

$$f = \delta \left(F_1 \left(\left[z^h, z^w \right] \right) \right) \tag{4}$$

$$g^h = \sigma \left(F_h \left(f^h \right) \right) \tag{5}$$

$$g^w = \sigma \left(F_w \left(f^w \right) \right) \tag{6}$$

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \tag{7}$$

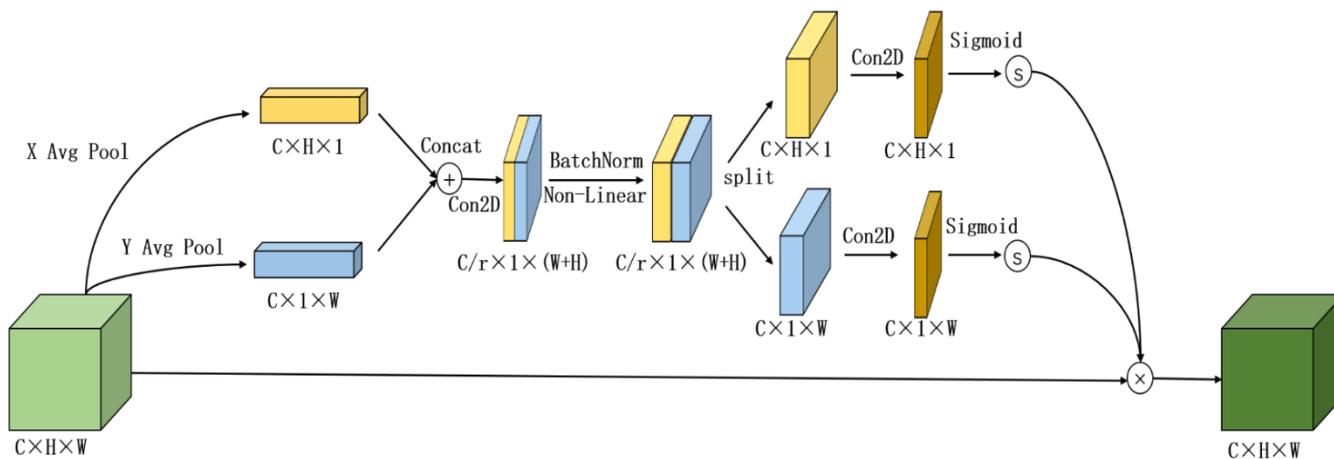


Figure 11. Structure of the coordinate attention mechanism.

4.2.2. Feature Fusion-Enhanced BiFPN

Feature pyramids emerged as a feature fusion framework that was derived to detect objects at different scales. However, the traditional structures of feature pyramids such as FPN (feature pyramid network) [34] and PANet add up different input features when fusing them. Since these different input features have different resolutions, the contributions of these features to the fused features usually vary. To solve this problem, we adopt BiFPN, a feature fusion framework, and replace the original PANet with BiFPN in the neck module. The main improvement made at this stage is to remove any node with only one input edge,

since without feature fusion, this will contribute less to the feature network with different feature fusion. Secondly, if the original input node is at the same level as the output node, the amount adds an extra edge between the input node and the output node to fuse more features without adding too much cost.

PANet simply adds up the different features when fusing them, as shown in Equations (8) and (9). Since the resolution of different input features is different, the contribution to the input features is usually also different. To solve this problem, BiFPN adjusts the contribution of each scale by adding weights to each scale feature. We take the features of the middle layer as an example, and describe the two fusion features of BiFPN at level 2, where P_2^{td} is the intermediate feature at level 2 of the top-down path and P_2^{out} is the output feature at level 2 of the bottom-up path, as shown in Equations (10) and (11). All of the other features are constructed similarly. It is worth noting that BiFPN uses a depthwise separable convolution for feature fusion and adds batch normalization and activation after each convolution. The principle of operation of PANet and BiFPN is illustrated in Figure 12.

$$P_2^{td} = Conv(P_2^{in} + Resize(P_3^{out})) \tag{8}$$

$$P_2^{out} = Conv(P_2^{td} + Resize(P_1^{out})) \tag{9}$$

$$P_2^{td} = Conv\left(\frac{\omega_1 P_2^{in} + \omega_2 Resize(P_3^{in})}{\omega_1 + \omega_2 + \epsilon}\right) \tag{10}$$

$$P_2^{out} = Conv\left(\frac{\omega'_1 P_2^{in} + \omega'_2 P_2^{td} + \omega'_3 Resize(P_1^{out})}{\omega'_1 + \omega'_2 + \omega'_3 + \epsilon}\right) \tag{11}$$

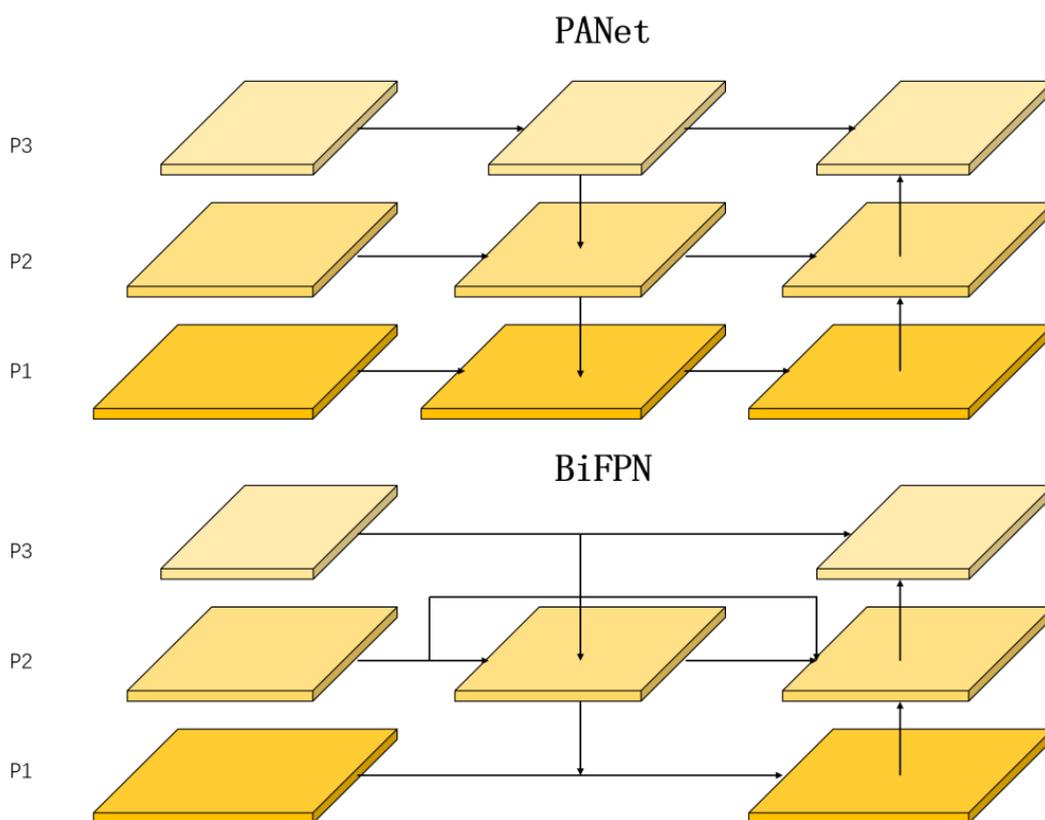


Figure 12. Diagram of the structure of PANet and BiFPN.

4.3. Proposed Framework

Figure 13 illustrates the framework of our transmission line defect detection method. The framework consists of data acquisition, pre-processing, and the BC-YOLO network testing module. We use remote sensing images taken during UAV power inspections to construct a dataset of vibration dampers and insulators called DVDI.

The specific steps in the transmission line defect detection process are as follows:

1. Remote sensing images of transmission lines are taken by the UAV during a power inspection.
2. The images are pre-processed using the gamma transform, and the defect dataset is expanded by rotational mirroring.
3. Labellmg is used to label the dataset, and the categories and boxes of insulators and vibration dampers are saved in an XML file.
4. The dataset is divided into a training set, validation set, and test set in the ratio of 6:2:2, and the resolution of the images is adjusted to 416×416 after feeding into the network.
5. The divided dataset is trained using BC-YOLO.
6. The loss function is observed during training, and the network weights are saved when the loss is minimized.
7. The saved network weights are used to detect insulators and vibration dampers with anomalies.

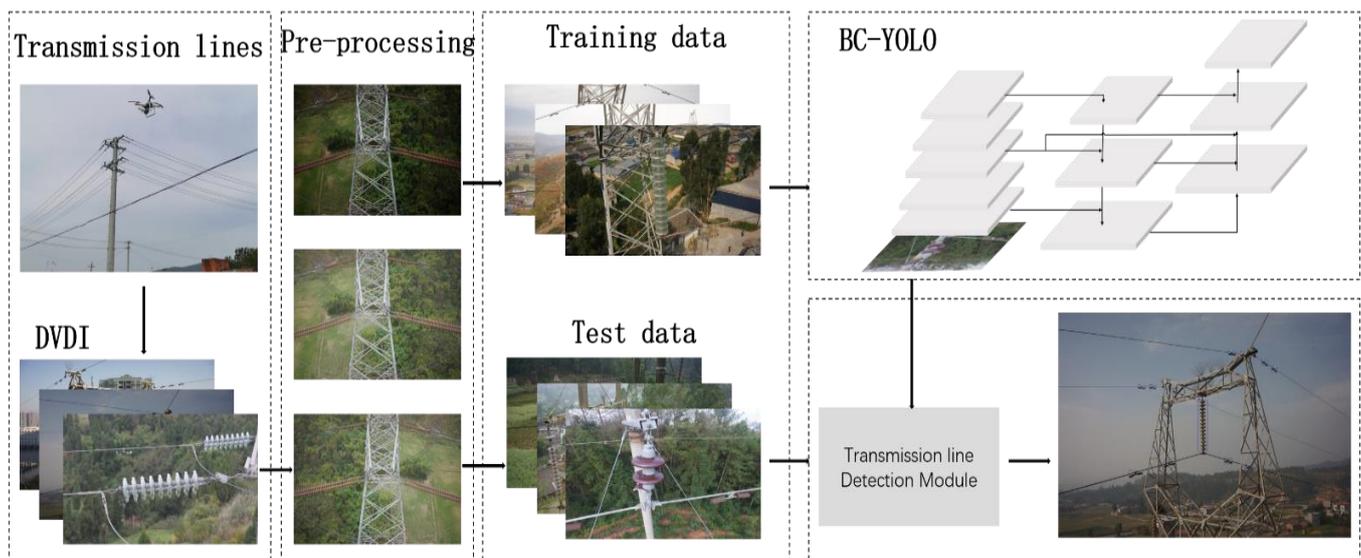


Figure 13. Schematic diagram of the proposed framework for insulator and vibration damper defect detection.

5. Experimental Results and Analysis

5.1. Experimental Environment and Parameters

The software environment and hardware parameters we used during the experiment are shown in Table 2.

Table 2. Software and hardware configuration of the experimental environment.

Platform	Configuration
Integrated development environment	PyCharm
Scripting language	Python3.8
Deep learning frame	PyTorch1.9.1
CPU model	Inter Core i7-9700k
Operating system	Ubuntu18.04 LTS 64-bits
GPU model	NVIDIA GeForce RTX 2080Ti
GPU accelerator	CUDA 10.2
Neural network accelerator	cuDNN7.6.5

The parameters of our experiments in training BC-YOLO are shown in Table 3.

Table 3. BC-YOLO experimental training parameters.

Parameter	Configuration
Neural network optimizer	SGD
Learning rate	0.01
Training epochs	300
Momentum	0.937
Batch size	4
Weight decay	0.0005

5.2. Performance Evaluation Index

In this paper, we use the Precision, Recall, Average Precision, and Mean Average Precision as evaluation indices for the network performance. The Recall rate represents how many true positive samples are retrieved by the network compared to the total number of positive instances, and is calculated as shown in Equation (12). Precision indicates how many of the predicted positive samples are positive, and the calculation method is shown in Equation (13). The AP is the area of the region enclosed by the Precision–Recall (P-R) curve and the coordinate axis, which is calculated as shown in Equation (14). The mAP is the average accuracy, which measures the overall detection effect of the network, as shown in Equation (15). The confusion matrix is given in Table 4.

Table 4. Confusion matrix.

	True	False
Positive	TP	FP
Negative	NT	FN

TP is predicted by the network as a positive sample of positive class, FP is predicted by the network as a negative sample of positive class, FN is predicted by the network as a positive sample of negative class, and TN is predicted by the network as a negative sample of negative class.

$$\text{Recall} = \frac{TP}{TP+FP} \quad (12)$$

$$\text{Precision} = \frac{TP}{TP+FN} \quad (13)$$

$$\text{AP} = \int_0^1 P(R)dR \quad (14)$$

$$\text{mAP} = \frac{\sum_{i=0}^{i=n} AP}{n} \quad (15)$$

5.3. Results and Analysis

5.3.1. Comparative Experiments on Attention Mechanisms

In this paper, we compare the three common attention mechanisms of SE, CBAM, and CA, which are commonly used in target detection tasks, and the results are shown in Table 4.

The three attention mechanisms are consistent in the positions they add in the YOLOv5 network. There are four versions of the YOLOv5 detection network, namely YOLOv5x, YOLOv5l, YOLOv5m, and YOLOv5s. The model used in this paper is YOLOv5x, which has the highest accuracy among the four versions. Firstly, adding the SE attention mechanism to the backbone improves the Recall by 1.9 points and the mAP by 1.9 points compared to the original YOLOv5x. Secondly, when the CBAM attention mechanism is added, the Recall is improved by 1.9 points and the mAP by 1.6 points. Finally, adding the CA attention mechanism improves the Recall by 4.1 points and the mAP by 2.0 points. From the results shown in the Table 5 below, we see that the effect of CA is the most significant.

Table 5. Effects of different attention mechanisms on network performance.

Baseline	SE	CBAM	CA	Recall(%)	Precision(%)	mAP(%)
YOLOv5x	✓			82.7	89.9	86.4
		✓		84.6	88.1	88.3
			✓	84.6	87.4	88.0
				86.8	88.0	88.4

We use heat maps to visualize the output feature maps for adding different attention mechanisms, as shown in Figure 14. From the results shown in the heat map of Figure 14, the network module with the CA mechanism detects the critical parts of transmission line detection more accurately than with SE or CBAM.

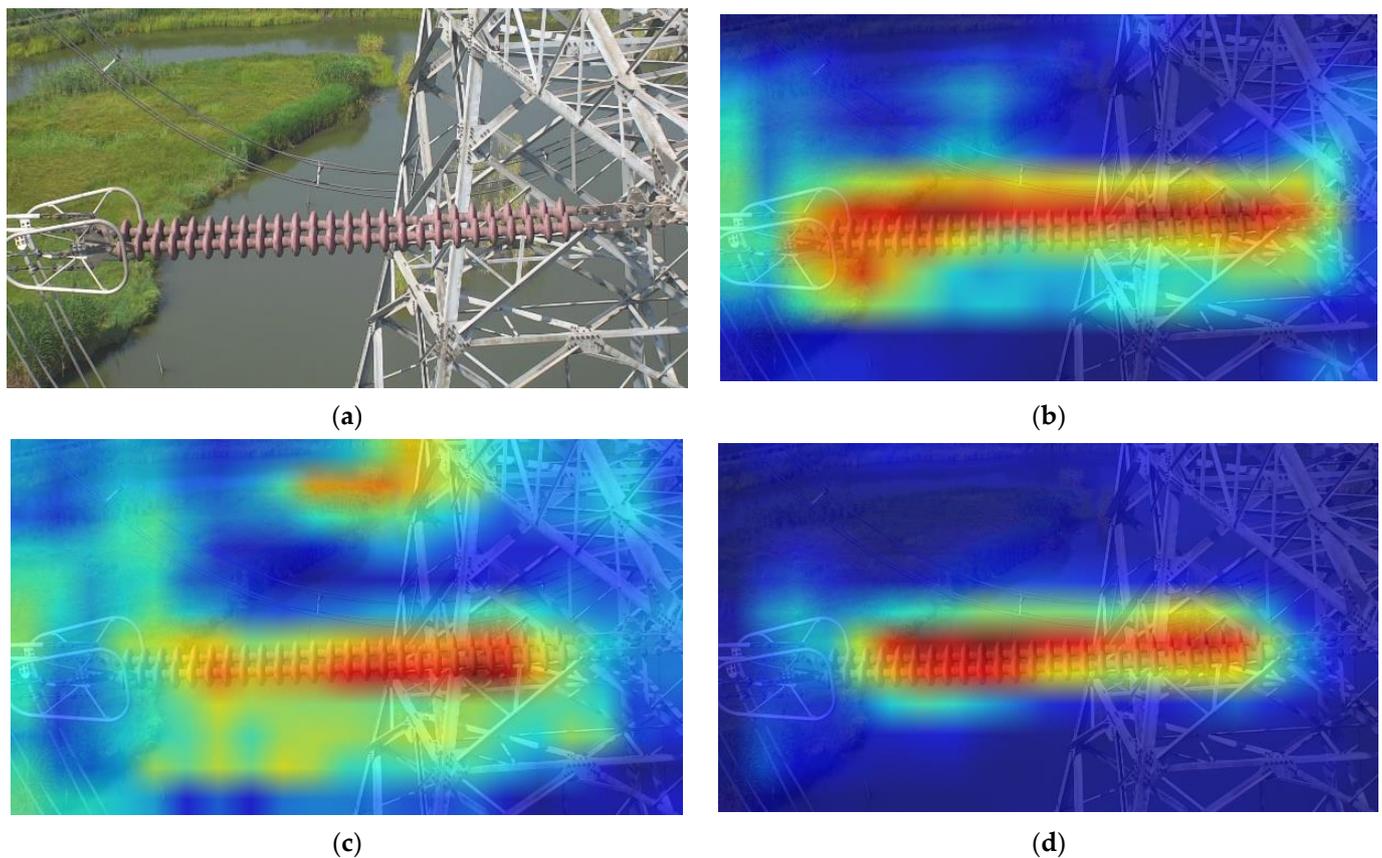


Figure 14. Heat maps for different attention modules: (a) original image; (b) SE; (c) CBAM; (d) CA.

5.3.2. Comparison of Experiments That Add Different Modules

In this experiment, we first used YOLOv5x as a baseline and tested the effect of adding the new feature pyramid BiFPN. The results showed that the mAP was improved by 1.8

points, and Recall by 2.7 points. This indicates that the new feature fusion can effectively fuse more feature layers. At the same time, in the process of feature fusion, the network pays more attention to the input features with great contribution, which enhances the learning ability of the model. In the second test, we compared the addition of the CA attention mechanism. The results showed that the mAP increased by two points and Recall by 4.1 points compared to the original baseline, which indicates that adding the attention mechanism to the backbone can effectively help the network to capture important information and long-distance dependencies when extracting features. Finally, the effect of fusing BiFPN and CA was explored, and the results showed that the mAP improved by 2.7 points and Recall by 4.0 points compared to the original baseline, meaning that the fusion of BiFPN and CA can effectively improve the accuracy of transmission line defect detection. The network detection results are presented in Table 6.

Table 6. Effect of each module on network performance.

Baseline	BiFPN	CA	Recall(%)	Precision(%)	mAP(%)
YOLOv5x			82.7	89.9	86.4
	✓		85.4	87.3	88.2
		✓	86.8	88.0	88.4
	✓	✓	86.7	86.2	89.1

5.3.3. Comparison of Different Object Detection Networks

To further verify the advantages of our method, we compare it with SSD [35], RetinaNet [36], CenterNet [37], and YOLOv4 [38] methods, and the detection results are shown in Table 7. Our method outperformed the other methods for all classes of mAP, especially for small targets such as vibration dampers. This is due to the small sizes of the numerous vibration damper targets, some of which are easily confused with forests, fields, and dead leaves due to their color, making it difficult for general target detection methods to detect them. In contrast, the methods used in the YOLO series can effectively handle detection tasks involving large differences in target sizes. When the CA mechanism and the new Bi-FPN feature fusion method were added, the accuracy of all categories improved significantly, indicating that this module can effectively improve the detection accuracy of the network. A visualization of the test results for each type of method is shown in Figure 15.

Table 7. Detection results for some mainstream object detection networks.

Method	1(AP)	2(AP)	3(AP)	4(AP)	mAP(%)
SSD	73.0	81.0	22.0	47.0	55.86
RetinaNet	78.0	78.0	26.0	36.0	54.37
CenterNet	85.0	85.0	52.0	59.0	70.33
YOLOv4	87.0	78.0	74.0	83.0	80.26
YOLOv5x	88.4	87.0	82.9	87.4	86.4
Proposed method	90.0	89.2	84.3	92.7	89.1

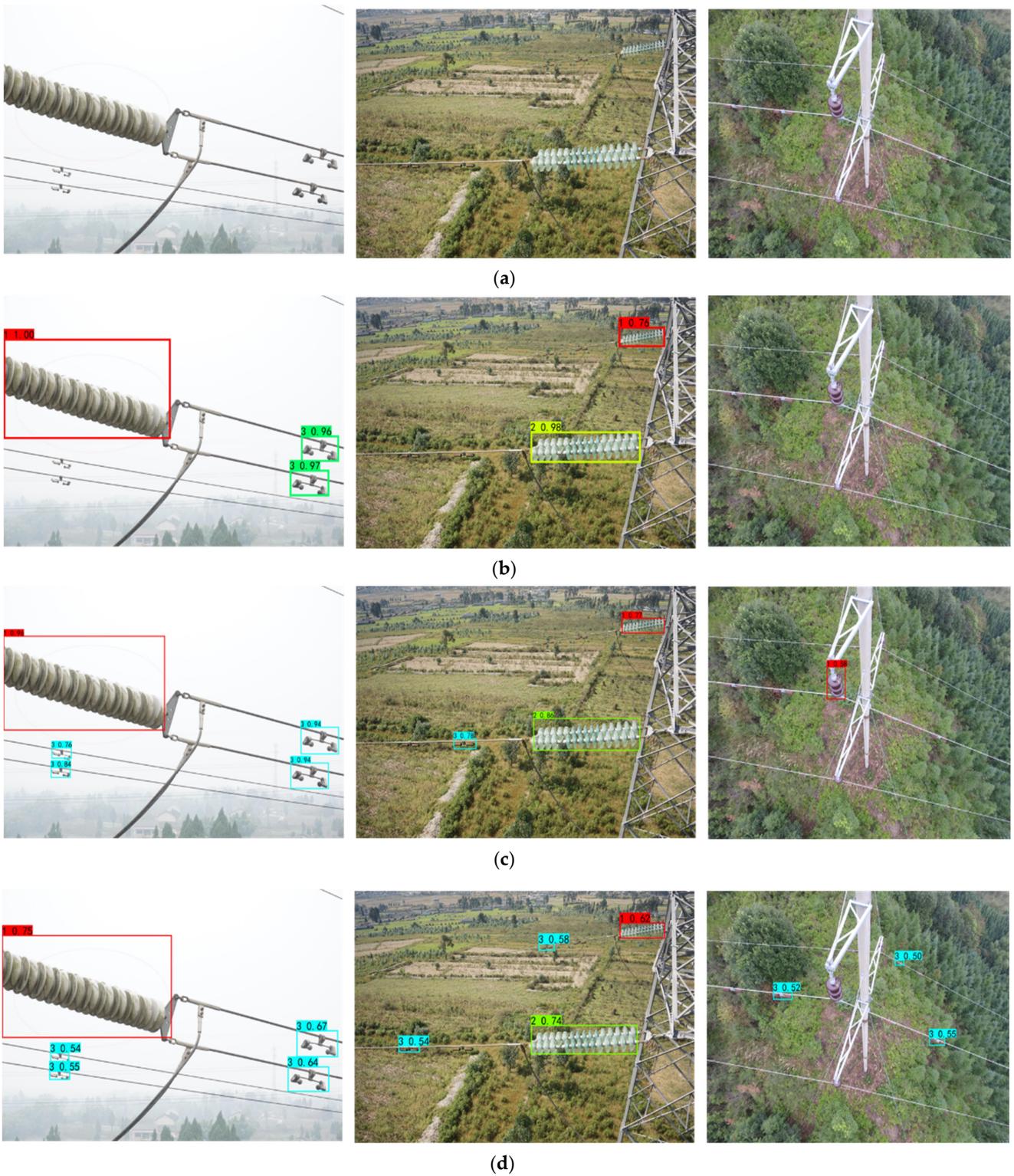


Figure 15. Cont.

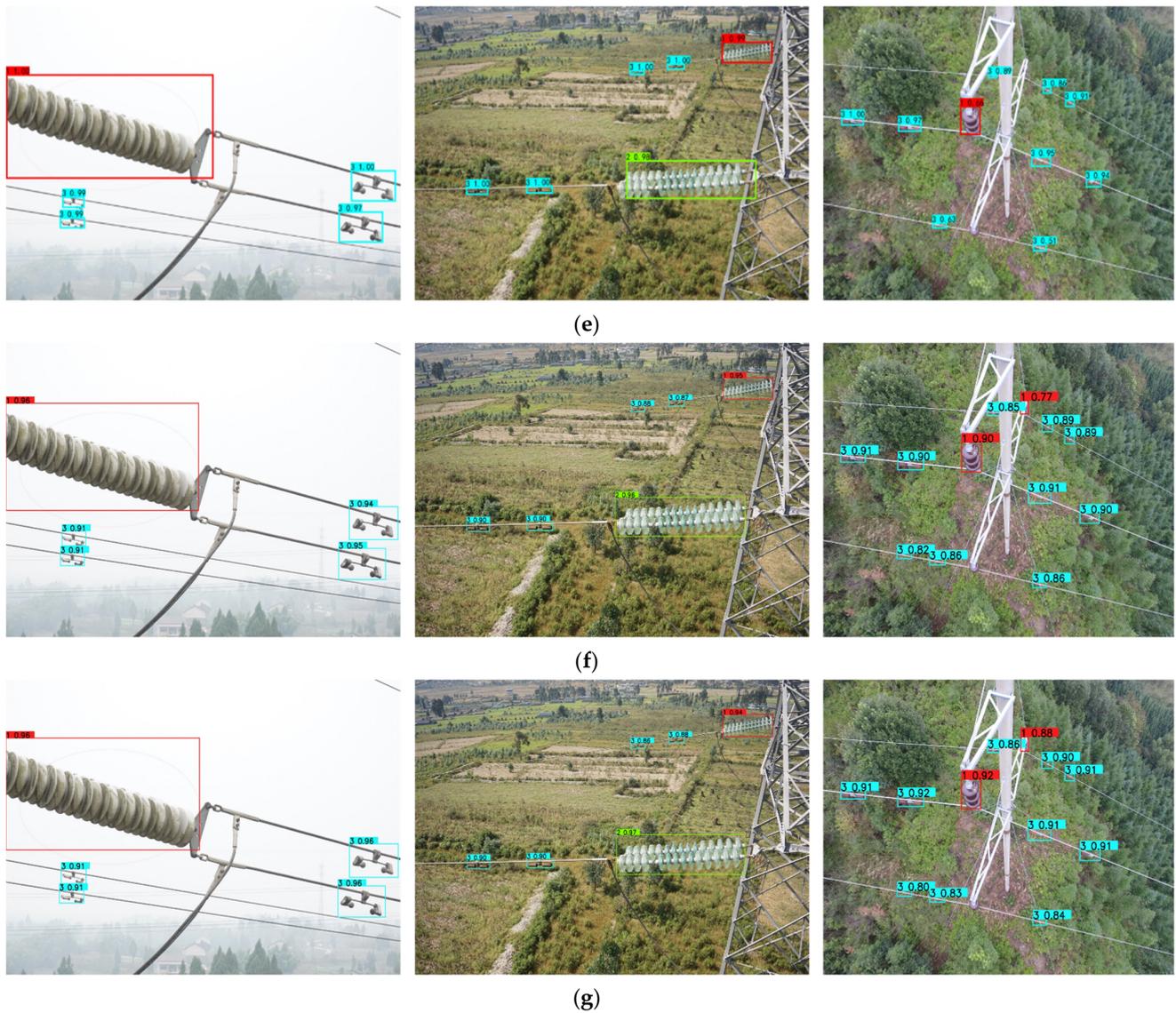


Figure 15. Results on the test dataset: (a) original images; (b) SSD; (c) RetinaNet; (d) CenterNet; (e) YOLOv4; (f) YOLOv5x; (g) proposed method.

6. Discussion

The confusion matrix is a standard form of representing the accuracy evaluation, also known as the error matrix, and is represented by an $n \times n$ matrix form. A confusion matrix was obtained from our training results after normalization, as shown in Figure 16, where each row represents the true attribution category of the data and each column represents the category predicted by the network. The total amount of data in each row represents the total amount of true data in that category, and the values in each column represent the number of categories predicted by the network for that category.

From the confusion matrix, it can be seen that the detection of defects in both insulators and vibration dampers is affected by the background; the vibration dampers are influenced more strongly by the background than the insulators, as the latter are larger and less likely to be obscured. The color of an insulator is less likely to be confused with the background color, and there is therefore less interference from the background. In contrast, the vibration dampers are easily obscured by transmission towers, transmission lines, and buildings due to their small size. The number of vibration dampers is also large, and since they are easily confused with the background color, this causes certain difficulties in detection. In particular, normal vibration dampers are easily confused with dead leaves, houses, and forests, and the background therefore has more influence on the normal vibration dampers than the defective ones. Due to the excessive number of insulator pieces in some insulator strings, their damage may not be easily detected and therefore may be mistaken for normal insulators.

Predicted	1	0.90	0.14			0.39
	2	0.03	0.81			0.08
	3			0.78	0.03	0.46
	4			0.01	0.87	0.07
	Background FN	0.06	0.04	0.21	0.10	
		1	2	3	4	Background FP
		True				

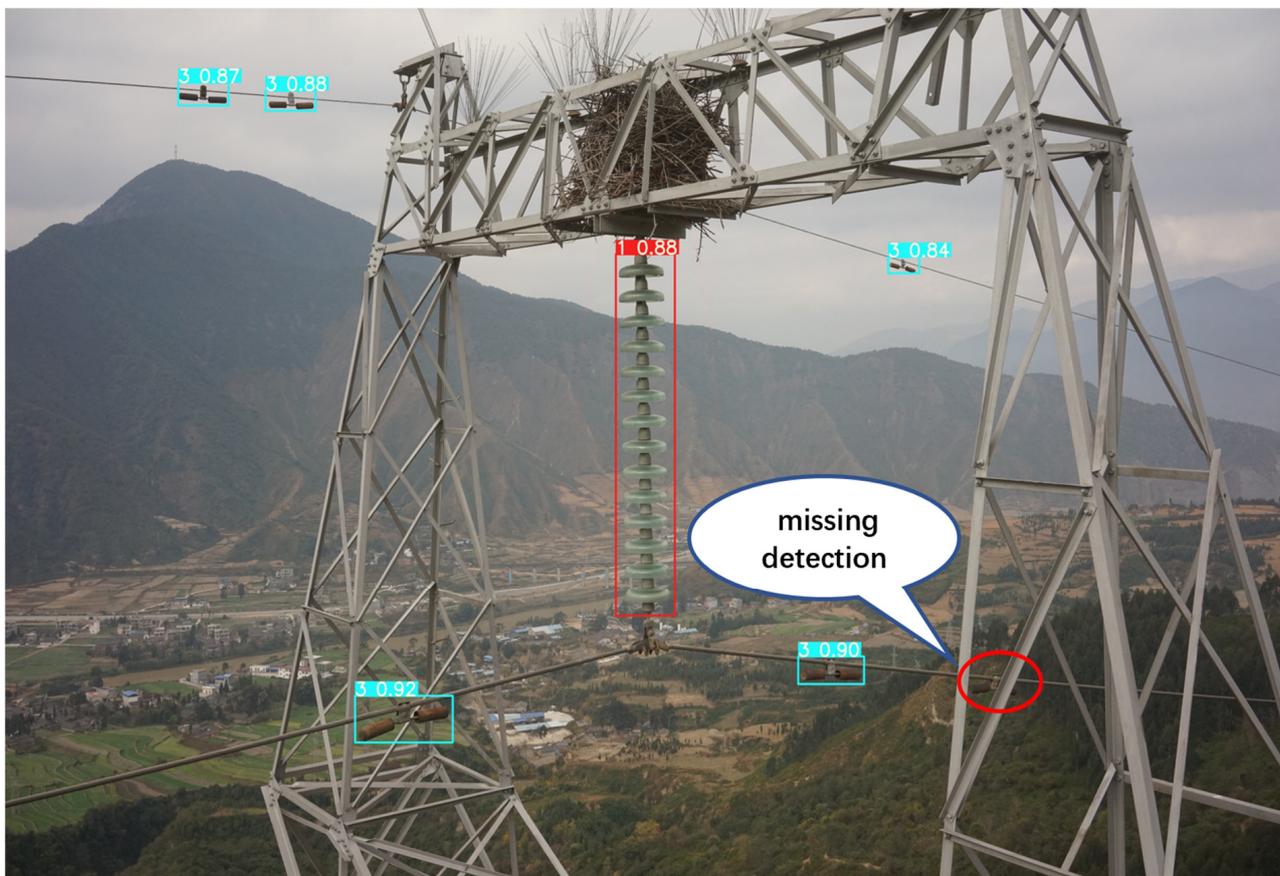
Figure 16. Confusion matrix.

Although the detection accuracy of defective vibration dampers and insulators images is improved using the method proposed in the paper, a few defective vibration dampers were still missed. Some examples are shown in Figure 17. In Figure 17a, the missed detection of the defective vibration dampers is caused by the fact that their color is very close to the color of the land. In Figure 17b, a defective vibration damper is partially occluded, so it is also undetected. Therefore, occlusion and similar color are the main reasons for missed detection, which are also the limitations of the proposed method.



(a)

Figure 17. Cont.



(b)

Figure 17. Example of partially missed detection: (a) Color confusion miss detection; (b) Missing detection from masking.

7. Conclusions

To accurately detect insulators and vibration dampers on transmission lines with a complex background, in low light, and with large size differences between targets, we have proposed an optimized version of YOLOv5 through the use of methods such as a CA mechanism and Bi-FPN enhanced feature fusion, to develop the BC-YOLO insulator and vibration damper defect detection method. The findings of this paper can be summarized in the following three points: (i) the datasets used in the experiments contained pictures of multiple types of insulators and vibration dampers, taken during the power inspection process; (ii) a CA mechanism is added to the feature extraction module of YOLOv5, which allows the mobile network to obtain information about a larger area without introducing a larger overhead by embedding the location information into the channel attention; (iii) for the case of insulators and vibration dampers with large differences in size, we introduce Bi-FPN, based on the original feature fusion method, which can effectively improve the efficiency of feature fusion and improve the detection accuracy of small targets. Through an experimental comparison, it was found that the mAP@0.5 for BC-YOLO reached 89.1% on the test set of the DVDI, a value 2.7% higher than for YOLOv5.

In future research, we will collect additional data on defective insulators and vibration dampers taken during UAV inspections, and will expand the types of defects for the insulators and vibration dampers, which will allow the network to be adapted to the detection of insulators and vibration dampers with various types of defects. We will also investigate how to compress the model as a way to increase the detection speed while maintaining the detection accuracy of the model.

Author Contributions: Conceptualization, X.D.; data curation, W.B. and N.W.; formal analysis, W.B. and X.D.; investigation, N.W. and X.Y.; methodology, X.D., W.B., and M.Y.; resources, W.B., N.W., and X.Y.; supervision, W.B. and N.W.; validation, X.D. and M.Y.; writing—original draft, X.D.; writing—review and editing, W.B., X.D., N.W., and M.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (Grant No. 2020YFF0303803) and Anhui Natural Science Foundation (Grant No. 2208085MC60).

Data Availability Statement: <https://github.com/Emp-8/DVDI>, accessed on 12 October 2022.

Acknowledgments: The authors sincerely thank the anonymous reviewers for their critical comments and suggestions for improving the manuscript.

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

References

1. Sadykova, D.; Pernebayeva, D.; Bagheri, M.; James, A. IN-YOLO: Real-time detection of outdoor high voltage insulators using UAV imaging. *IEEE Trans. Power Deliv.* **2019**, *35*, 1599–1601. [CrossRef]
2. Liang, H.; Zuo, C.; Wei, W. Detection and evaluation method of transmission line defects based on deep learning. *IEEE Access* **2020**, *8*, 38448–38458. [CrossRef]
3. Wang, S.; Liu, Y.; Qing, Y.; Wang, C.; Lan, T.; Yao, R. Detection of Insulator Defects With Improved ResNeSt and Region Proposal Network. *IEEE Access* **2020**, *8*, 184841–184850. [CrossRef]
4. Siddiqui, Z.A.; Park, U. A drone based transmission line components inspection system with deep learning technique. *Energies* **2020**, *13*, 3348. [CrossRef]
5. Jiang, S.; Jiang, W.; Huang, W.; Yang, L. UAV-based oblique photogrammetry for outdoor data acquisition and offsite visual inspection of transmission line. *Remote Sens.* **2017**, *9*, 278. [CrossRef]
6. Liu, X.; Lin, Y.; Jiang, H.; Miao, X.; Chen, J. Slippage fault diagnosis of dampers for transmission lines based on faster R-CNN and distance constraint. *Electr. Power Syst. Res.* **2021**, *199*, 107449. [CrossRef]
7. Yang, L.; Fan, J.; Liu, Y.; Li, E.; Peng, J.; Liang, Z. A review on state-of-the-art power line inspection techniques. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 9350–9365. [CrossRef]
8. Katrasnik, J.; Pernus, F.; Likar, B. A survey of mobile robots for distribution power line inspection. *IEEE Trans. Power Deliv.* **2009**, *25*, 485–493. [CrossRef]
9. Nguyen, V.N.; Jenssen, R.; Roverso, D. Automatic autonomous vision-based power line inspection: A review of current status and the potential role of deep learning. *Int. J. Electr. Power Energy Syst.* **2018**, *99*, 107–120. [CrossRef]
10. Matikainen, L.; Lehtomäki, M.; Ahokas, E.; Hyypä, J.; Karjalainen, M.; Jaakkola, A.; Kukko, A.; Heinonen, T. Remote sensing methods for power line corridor surveys. *ISPRS J. Photogramm. Remote Sens.* **2016**, *119*, 10–31.
11. Tiantian, Y.; Guodong, Y.; Junzhi, Y. Feature fusion based insulator detection for aerial inspection. In Proceedings of the 2017 36th Chinese Control Conference (CCC), Dalian, China, 26–28 July 2017; pp. 10972–10977.
12. Reddy, M.J.B.; Chandra, B.K.; Mohanta, D.K. A DOST based approach for the condition monitoring of 11 kV distribution line insulators. In *IEEE Transactions on Dielectrics and Electrical Insulation*; IEEE: Piscataway, NJ, USA, 2011; Volume 18, pp. 588–595. [CrossRef]
13. Reddy, M.J.B.; B, K.C.; Mohanta, D.K. Condition monitoring of 11 kV distribution system insulators incorporating complex imagery using combined DOST-SVM approach. In *IEEE Transactions on Dielectrics and Electrical Insulation*; IEEE: Piscataway, NJ, USA, 2013; Volume 20, pp. 664–674. [CrossRef]
14. Li, B.; Wu, D.; Cong, Y.; Xia, Y.; Tang, Y. A method of insulator detection from video sequence. In Proceedings of the 4th International Symposium on Information Science and Engineering (ISISE), Shanghai, China, 14–16 December 2012; pp. 386–389.
15. Zhao, Z.; Liu, N.; Wang, L. Localization of multiple insulators by orientation angle detection and binary shape prior knowledge. In *IEEE Transactions on Dielectrics and Electrical Insulation*; IEEE: Piscataway, NJ, USA, 2015; Volume 22, pp. 3421–3428. [CrossRef]
16. Wu, Q.; An, J.; Lin, B. A Texture Segmentation Algorithm Based on PCA and Global Minimization Active Contour Model for Aerial Insulator Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *5*, 1509–1518. [CrossRef]
17. Liao, S.; An, J. A robust insulator detection algorithm based on local features and spatial orders for aerial images. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 963–967. [CrossRef]
18. Jabid, T.; Uddin, M.Z. Rotation invariant power line insulator detection using local directional pattern and support vector machine. In Proceedings of the International Conference on Engineering and Technology Innovations 2016 (ICISSET), Wuhan, China, 26–24 March 2016; pp. 1–4.
19. Zhao, Z.; Xu, G.; Qi, Y. Representation of binary feature pooling for detection of insulator strings in infrared images. In *IEEE Transactions on Dielectrics and Electrical Insulation*; IEEE: Piscataway, NJ, USA, 2016; Volume 23, pp. 2858–2866. [CrossRef]

20. Guifeng, Z.; Zhiren, T. Swarm optimization clustering for image segmentation of insulators. In Proceedings of the IEEE Conference on Industrial Electronics and Applications (ICIEA), Xi'an, China, 16–21 June 2019; pp. 397–401.
21. Li, F.; Xin, J.; Chen, T.; Xin, L.; Wei, Z.; Li, Y.; Zhang, Y.; Jin, H.; Tu, Y.; Zhou, X.; et al. An Automatic Detection Method of Bird's Nest on Transmission Line Tower Based on Faster_RCNN. *IEEE Access* **2020**, *8*, 164214–164221. [[CrossRef](#)]
22. Wanguo, W.; Zhenli, W.; Bin, L.; Yuechen, Y.; Xiaobin, S. Typical Defect Detection Technology of Transmission Line Based on Deep Learning. In Proceedings of the 2019 Chinese Automation Congress (CAC), Hangzhou, China, 22–24 November 2019; pp. 1185–1189. [[CrossRef](#)]
23. Wu, C.; Ma, X.; Kong, X.; Zhu, H. Research on insulator defect detection algorithm of transmission line based on CenterNet. *PLoS ONE* **2021**, *16*, e0255135. [[CrossRef](#)]
24. Wu, W.; Li, Q. Machine Vision Inspection of Electrical Connectors Based on Improved Yolo v3. *IEEE Access* **2020**, *8*, 166184–166196. [[CrossRef](#)]
25. Bao, W.; Ren, Y.; Wang, N.; Hu, G.; Yang, X. Detection of Abnormal Vibration Dampers on Transmission Lines in UAV Remote Sensing Images with PMA-YOLO. *Remote Sens.* **2021**, *13*, 4134. [[CrossRef](#)]
26. Liu, B.; Wang, X.; Jin, W.; Chen, Y.; Liu, C. Infrared image detail enhancement based on local adaptive gamma correction. *Chin. Opt. Lett.* **2012**, *10*, 25–29.
27. Everingham, M.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal Visual Object Classes (VOC) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [[CrossRef](#)]
28. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [[CrossRef](#)]
29. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768. [[CrossRef](#)]
30. Hou, Q.; Zhou, D.; Feng, J. Coordinate Attention for Efficient Mobile Network Design. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 13708–13717. [[CrossRef](#)]
31. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and Efficient Object Detection. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 10778–10787. [[CrossRef](#)]
32. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*; IEEE: Piscataway, NJ, USA, 2020; Volume 42, pp. 2011–2023. [[CrossRef](#)]
33. Woo, S.; Park, J.; Lee, J.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the 2018 European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19. [[CrossRef](#)]
34. Lin, T.Y.; Dollar, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. [[CrossRef](#)]
35. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. *arXiv* **2015**, arXiv:1512.02325.
36. Lin, T.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2999–3007. [[CrossRef](#)]
37. Zhou, X.Y.; Wang, D.Q.; Krhenbühl, P. Objects as points. *arXiv* **2019**, arXiv:1904.07850.
38. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-J.M. YOLOv4 optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.