Article

Insulator Breakage Detection Based on Improved YOLOv5

Gujing Han [1,†], Min He [1,†], Mengze Gao [1,*], Jinyun Yu [2,†], Kaipei Liu [2,†] and Liang Qin [2,†]

[1] School of Electronic and Electrical Engineering, Wuhan Textile University, Wuhan 430200, China; gjhan@wtu.edu.cn (G.H.); 13098803217@163.com (M.H.)
[2] School of Electrical and Automation, Wuhan University, Wuhan 430072, China; yujinyun0707@126.com (J.Y.); kpliu@whu.edu.cn (K.L.); qinliang@whu.edu.cn (L.Q.)
* Correspondence: 2015053026@mail.wtu.edu.cn; Tel.:+86-1582-753-4392
† These authors contributed equally to this work.

Abstract: Aerial images have complex backgrounds, small targets, and overlapping targets, resulting in low accuracy of intelligent detection of overhead line insulators. This paper proposes an improved algorithm for insulator breakage detection based on YOLOv5: The ECA-Net (Efficient Channel Attention Network) attention mechanism is integrated into its backbone feature extraction layer, and the effective distinction between background and target is achieved by increasing the weight of important channels. A bidirectional feature pyramid network is added to the feature fusion layer, and large-scale images with more original information are combined to effectively retain small target features. Incorporating a flexible detection frame selection algorithm Soft-NMS (Soft Non-Maximum Suppression) into the prediction layer to re-screen the target frame, thereby reducing the probability of mistaken deletion of overlapping targets. The effectiveness of the improved YOLOv5 algorithm is verified in the actual aerial image dataset, and the results show that the mean Average Precision (mAP) of the improved algorithm is 95.02% and the detection speed FPS (Frames Per Second) can reach 49.4 frames/s, which meets the real-time and accuracy requirements of engineering applications.

Keywords: insulator; small target detection; YOLOv5; Bi-FPN; ECA-Net; overlapping target detection

check for updates

Citation: Han, G.; He, M.; Gao, M.; Yu, J.;Liu, K.; Qin, L. Insulator Breakage Detection Based on Improved YOLOv5. *Sustainability* **2022**, *14*, 6066. https://doi.org/10.3390/su14106066

Academic Editor: J. C. Hernandez

Received: 21 April 2022
Accepted: 13 May 2022
Published: 17 May 2022

1. Introduction

The reliable operation of insulators is an important guarantee for the safety of high-voltage transmission lines and even power grids. Due to long-term exposure to the complex natural environment, the following faults are prone to exist, such as cracks in insulators caused by external forces, lightning overvoltage leading to insulator flashover and burns or even explosions, and fouling flash leading to cracks on the insulator surface. This paper focuses on the insulator breakage problem. The current mainstream inspection method for detecting insulators is to use UAVs (Unmanned Aerial Vehicles) to take pictures and perform damage identification, which substantially reduces the workload of maintenance personnel in field operations. However, the phenomenon of complex natural backgrounds, a small percentage of targets and overlapping targets in the aerial photography images of UAVs poses a great obstacle to the subsequent intelligent detection of insulator defects.

Traditional insulator image detection methods usually require manual intervention in the feature extraction link, such as the use of Hough transform linear detection, edge detection, watershed algorithm and the use of spatial and color information [1–7], which have weak adaptability, poor generalization ability and limited room for improvement in detection accuracy, and are mainly used for large target images with simple backgrounds, which are difficult to meet the needs of target detection in complex image backgrounds.

In recent years, deep learning has gained wide application in target detection in power systems, which can achieve precise location and fast classification of targets in images [8–14]. Deep learning target detection algorithms are divided into two categories, One-stage and Two-stage. Compared with Two-stage algorithms, One-stage omits the RPN (Region Proposal Network) candidate region generation, which significantly reduces the computational effort and directly uses the output of the convolutional neural network to characterize the detection results, and the YOLO (You Only Look Once) series algorithm is a typical representative of One-stage, featuring fast detection speed and high detection accuracy.

In the detection of targets with complex backgrounds, literature [15] proposed to add Focus structure in the input layer and combine it with YOLOv3 [16] target detection network to effectively detect small airplanes in complex backgrounds. Focus structure, mainly through matrix transformation, achieves data dimensionality reduction while completely retaining the original information, but YOLOv3 has the problems of inaccurate extraction of position information and low recall rate, especially in obscured and crowded scenes with low detection accuracy. Literature [17] improves the image clarity by SR-CNN (Super Resolution) network [18], and then uses the YOLOv4 algorithm [19] to improve the detection accuracy of glass insulators in complex backgrounds, but it is difficult to achieve end-to-end real-time detection. The literature [20] combined YOLOv2 [21] with CapsNet [22] to retain the insulator angle and direction and other feature information, which can identify insulator breakage in complex environments, but the model has more redundant information to increase the computational burden. The literature [23] used SENet (Squeeze and Excitation Network) attention mechanism [24] to improve the accuracy of the SSD (Single Shot Multi-Box Detector) network [25] for remote sensing building detection in complex backgrounds, but the computational effort of the fully connected structure used in the SENet attention mechanism for channel compression increases sharply with the increase of feature dimensionality.

In terms of small target detection, literature [26] combined cavity convolution in FPN (Feature Pyramid Network) [27] to increase the perceptual field and improve the small target detection accuracy of Faster R-CNN [28] for UAVs aerial photography. Literature [29] replaced the backbone of Faster R-CNN with DenseNet [30] and combined the FPN feature pyramid structure to improve its ability of small target detection, but the network structure based on Faster R-CNN has the problems of complex model and large computation. Meanwhile, literature [31] first used the SSD algorithm to locate the connection part of the defective bolt, and then used YOLOv3 to locate the defective bolt to achieve the detection of small targets, but the whole computation process is cumbersome and requires multiple models to locate. Literature [32] used an SSD network with residual structure, combined with a SENet attention mechanism to improve the detection of small targets of pin defects in transmission lines by models, but the lack of multi-scale feature fusion structure improves the detection speed. However, the detection accuracy will be subsequently reduced, especially not applicable to the detection of aerial images with multi-scale target characteristics.

In terms of overlapping target detection, literature [33] used the YOLOv4 model with ASPP (Atrous Spatial Pyramid Pooling) [34] structure to improve the overlap detection rate of hazardous materials under X-ray, although ASPP can be used to increase the perceptual field of the convolutional neural network. As the void rate increases near the feature map size, its ability to capture the full image context ability degrades to the role of an ordinary 1×1 filter, which is not conducive to feature extraction. The literature [35] improves the detection of overlapping targets in SSD networks by fusing manually extracted features: HOG features [36], RGB features, LBP features, and convolutional features, and combining them with NMS algorithms [37], which have a tedious feature extraction process. Furthermore, [38] uses the Cascade R-CNN algorithm [39] that uses a cascade of NMS values to select multiple candidate frames to locate overlapping insulators, comparing the work of selecting accurate candidate frames to reduce the detection speed of the model.

In 2020, Ultralytics introduced YOLOv5 [40], which combines the Focus structure, GIoU loss [41], and feature pyramid structure with the potential to handle complex backgrounds, small targets, and overlapping target problems in general images. Due to the high similarity between the background of insulators in aerial images such as forests, houses, etc., and glass and ceramic insulators, the smaller percentage of broken insulators in the images, and limited by the shooting distance, the proportion of broken insulators in the image is much smaller. Thus, the insulators photographed will overlap due to the angle of aerial photography. The current results of YOLOv5 directly applied to insulator detection do not meet the engineering requirements.

In summary, this paper uses YOLOv5 as the base model for insulator breakage detection of aerial images, and mainly makes three important improvements:

1. Fuses the attention mechanism ECA-Net [42] in its backbone feature extraction layer to compensate for the lack of information between channels by enhancing the information interaction between each channel and adaptively assigning the weights of background and target features;
2. Increases the proportion of small target feature maps in the feature fusion layer of the network through a two-way feature fusion network; the proportion of small target feature maps in the network is increased to effectively prevent the loss of small target information to detect small targets;
3. The Soft-NMS algorithm [43], which uses a reassignment of the scores of the original candidate frames to prevent overlapping candidate frames from being rejected, improves the detection accuracy of overlapping insulators.

2. Yolov5 Model

2.1. Principle of YOLOv5 Algorithm

The YOLOv5 algorithm consists of three parts: feature extraction layer, feature fusion layer, and prediction layer, as shown in Figure 1.

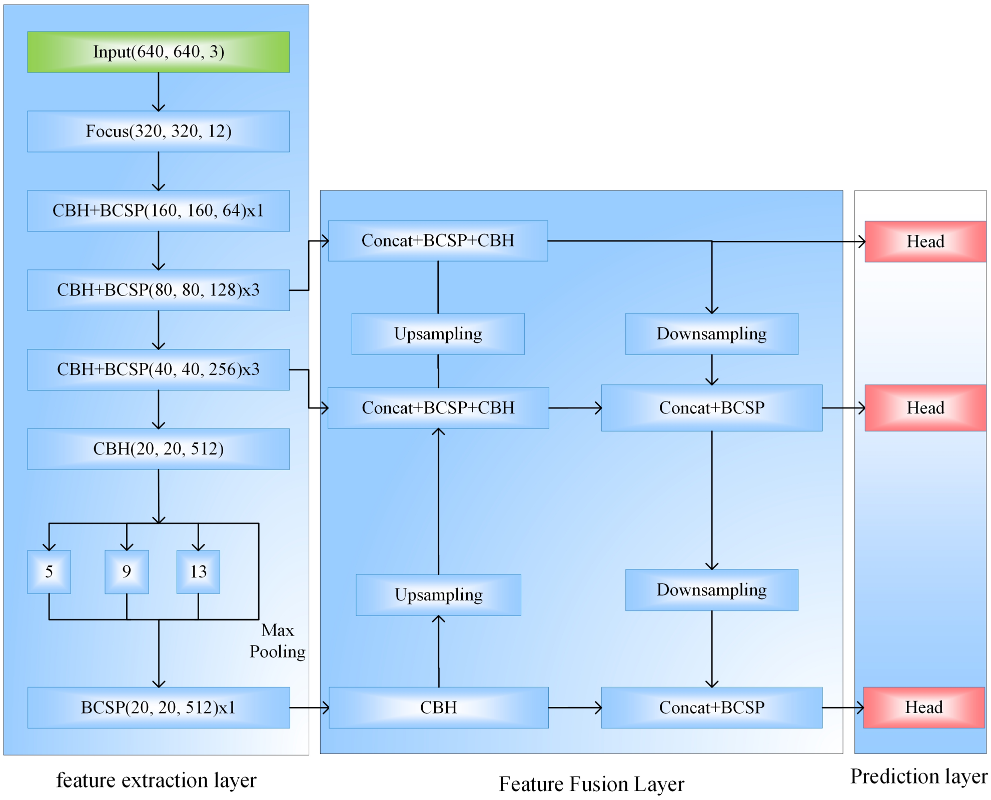


Figure 1. YOLOv5 model diagram.

2.1.1. Principle of YOLOv5 Algorithm

The feature extraction layer of YOLOv5 mainly uses Bottleneck-CSPNet as the backbone feature extraction structure. In Figure 1, CBH is composed of convolution, batch normalization, and activation function, where HardSwish is used for the activation function, which is calculated as Equations (1) and (2).

$$ReLU6 = min(6, max(0, x)) \tag{1}$$

$$HardSwish(x) = x\frac{ReLU6(x+3)}{6} \tag{2}$$

Compared with the LeakReLU of YOLOv3 and the Mish function of YOLOv4, HardSwish is numerically stable and computationally fast [44]. BCSP is the core structure of Bottleneck-CSPNet, as shown in Figure 2. The input features are divided into two branches on the left and right: the left side is a residual edge that maximizes the retention of the original information; the right main part carries out the residual stacking under the bottleneck structure, which can reduce the computation and also facilitate the model depth in deepening while its training process can converge better.

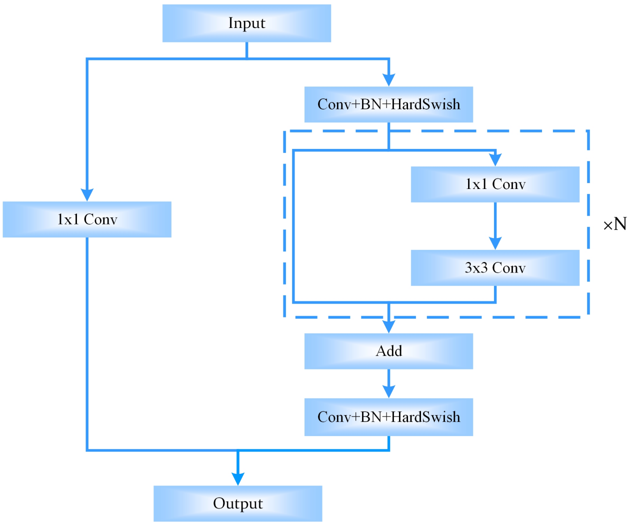


Figure 2. Structure of Bottleneck-CSPNet.

2.1.2. Feature Fusion Layer

The feature fusion network is shown in Figure 3, and its network is PANet (Path Aggregation-Network) [45]. The input image is first convolved and downsampled several times to get three scales (80, 40, 20) of features, and the features are subjected to bottom-up and top-down convolution and feature fusion processes to get three new outputs. The output features are continuously fused and stacked to improve the detection capability of targets at different scales.

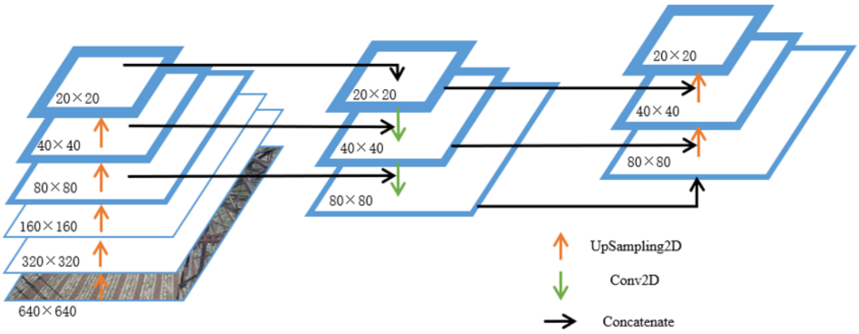


Figure 3. Structure of PANet.

2.1.3. Prediction Layer

The prediction layer maps the output three scale feature maps to the original image, locates targets at different scales by the different scale candidate frames obtained by the adaptive k-means clustering algorithm, and predicts the classification results.

2.1.4. Loss Function

The loss function of the YOLOv5 series contains three parts.

(1) The error loss between the prediction frame and the true frame is evaluated by GIoU (Generalized Intersection over Union) loss, as in Equation (3) and Figure 4.

$$\mathrm{GIoU}_{loss}(A,B) = 1 - (\mathrm{IoU}(A,B) - \frac{|C-(A\cap B)|}{|C|}) \tag{3}$$

where A and B represent the real frame and the predicted frame, respectively, C represents the smallest peripheral rectangle containing A and B frames, and $A \cap B$ is the overlapping area of the real frame and the predicted frame.

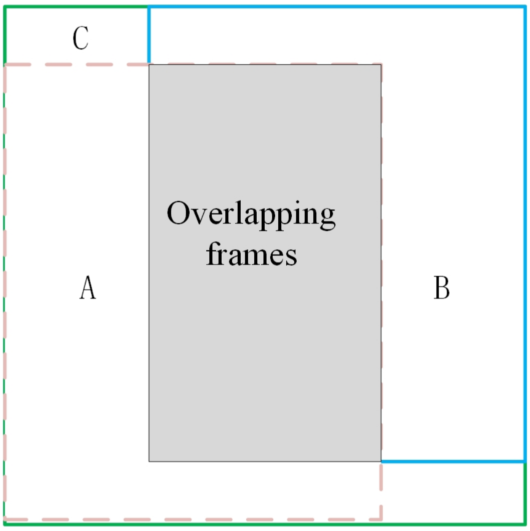


Figure 4. GIoU diagram.

(2) The loss of whether there is a target (0 for no target and 1 for target) is represented by the binary cross-entropy as in Equation (4).

$$L_{obj} = -[ylogx + (1-y)log(1-x)] \tag{4}$$

L_{obj} is the objective loss, where y is the true label, which takes the value of 1, and x is the predicted label, which takes the value of 0 or 1.

(3) The L_{cls} (classification loss) and L_{obj} (objective loss) are the same, just like Equation (3), but x is the number of predicted labels taking values between [0, 1]. The L_{cls} is used to calculate the loss into which category the target is classified, consists of a binary cross-entropy loss function. For a single feature map, the total loss is as in Equation (5).

$$L_{loss} = \mathrm{GIoU}_{loss} + L_{obj} + L_{cls} \tag{5}$$

Since there are three scales of the output feature maps, in general, large feature maps contain more information and are given a larger percentage of loss, and small feature maps contain less information and are assigned a smaller loss share, resulting in a total loss of Equation (6).

$$T_{loss} = 4 \times LOSS_{80\times80} + LOSS_{40\times40} + 0.25 \times LOSS_{20\times20} \tag{6}$$

2.1.5. Comparison of YOLOv5 and Each Model

Several basic algorithms of the YOLO series were compared in terms of algorithm complexity, detection speed of a single image, and detection speed of a video, and the results are shown in Table 1. The hardware environment tested was a local computer NVIDIA GeForce RTX 2060 SUPER 8G.

Table 1. Comparison of each base algorithm of YOLO.

Algorithm Model	Model Size (MB)	Video Detection Speed (FPS)	FLOPs (G)
YOLOv3	235	25.00	66.096
YOLOv4	244	22.21	60.334
YOLOv5s	27.8	68.18	17.060
YOLOv5m	83.2	40.62	51.427
YOLOv5l	182	25.12	115.603
YOLOv5x	340	10.00	219.026

The complexity of the algorithm is reflected by the FLOPs of the algorithm model. With the same software and hardware environment, larger FLOPs indicate a more complex model. In general, the lower the complexity of the algorithm, the faster the model can converge during training, which is beneficial for the embedded application of the model in mobile, while for video detection, speed is performed on the local computer graphics card NVIDIA GeForce RTX 2060 SUPER. As can be seen from Table 1, YOLOv5s algorithm complexity is the lowest, which is more than 83% lower than YOLOv3 and YOLOv4, and is only 27.8MB; the video detection speed is improved by more than 180%. Therefore, this paper improves and optimizes the algorithm based on YOLOv5s.

3. Improved Algorithm Based on YOLOv5

3.1. Backbone Feature Extraction Based on Attention Mechanism ECA-Net

The attention mechanism stems from the fact that humans selectively focus on the important parts of the received information. In mathematical language, this means that a set of weight coefficients is learned by the model autonomously, and this series of weights is assigned to each region of the input features so that the target information is weighted more and the irrelevant information is weighted less to achieve attention to the target.

As Figure 5 shows, the SENet network structure enhances channel relevance by downscaling and upscaling through full connectivity after global pooling. The fully connected layer is mainly calculated by combining the global feature maps with weight matrices, which is not conducive to the differentiation of complex background and target features. ECA-Net is an improved channel attention model based on SENet, which uses a one-dimensional convolution of convolution kernel size k instead of the fully connected layer for channel weighting of k proximity ranges to achieve local crossover and channel interaction, thus enhancing the network's attention to the local feature [45]. The ECA-Net network structure is shown in Figure 6.

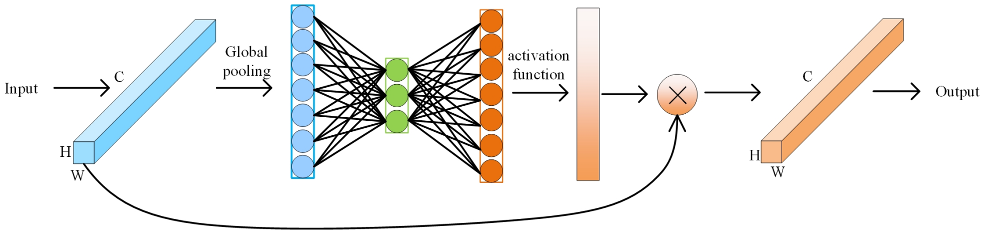


Figure 5. Structure of SENet.

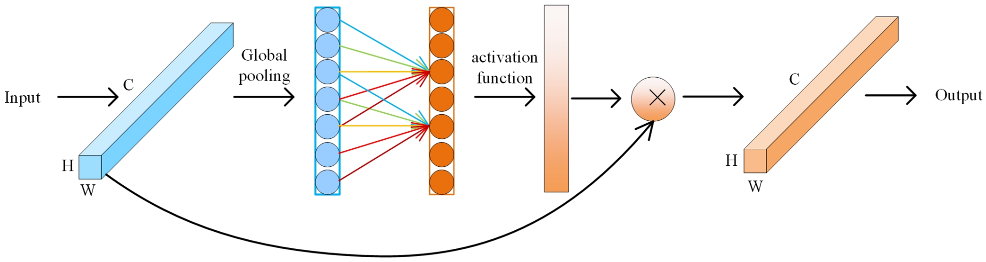


Figure 6. Structure of ECA-Net.

The k-value is calculated as in Equation (7), where channels is the number of channels of the input features. In Figure 6, every five channels are output as one channel, and padding is used to maintain a constant number of channels.

$$K = \frac{log_2(\text{channels}) + 1}{2} \tag{7}$$

3.2. Bi-Fpn-Based Feature Fusion Network

In order to avoid the problem of information loss of small targets in multi-layer convolutional changes as the feature map becomes smaller, the fusion of multi-scale features can avoid the loss of small targets to some extent. Meanwhile, one-way feature fusion structures such as FPN and PANet only consider the connection between the current layer and the adjacent layers, which makes it still easy to lose the original feature information during the convolutional feature extraction process, especially in the final output layer.

The feature fusion layer of YOLOv5 is replaced with a bi-directional feature fusion network Bi-FPN [46], which considers not only the current layer features and the adjacent layer features but also the original input layer features [47]. The structure is shown in Figure 7, and the loss of small targets is prevented by fusing large-scale feature maps with more information about small targets. Furthermore, to reduce the computational effort, Bi-FPN censors the top and bottom layer features in the middle structure.

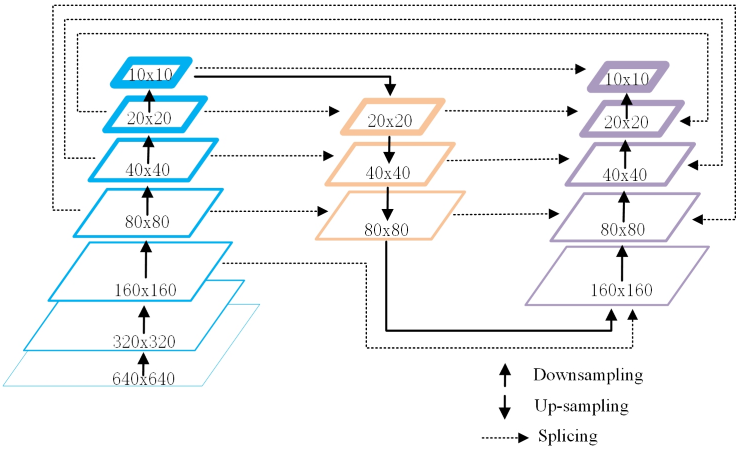


Figure 7. Structure of Bi-FPN.

3.3. Soft-NMS-Based Candidate Frame Algorithm

After the candidate frames are obtained, they are sent to NMS (Non-Maximum Suppression), which searches for local maxima and suppresses non-maxima to determine the final prediction frame. The flow is shown in Figure 8.

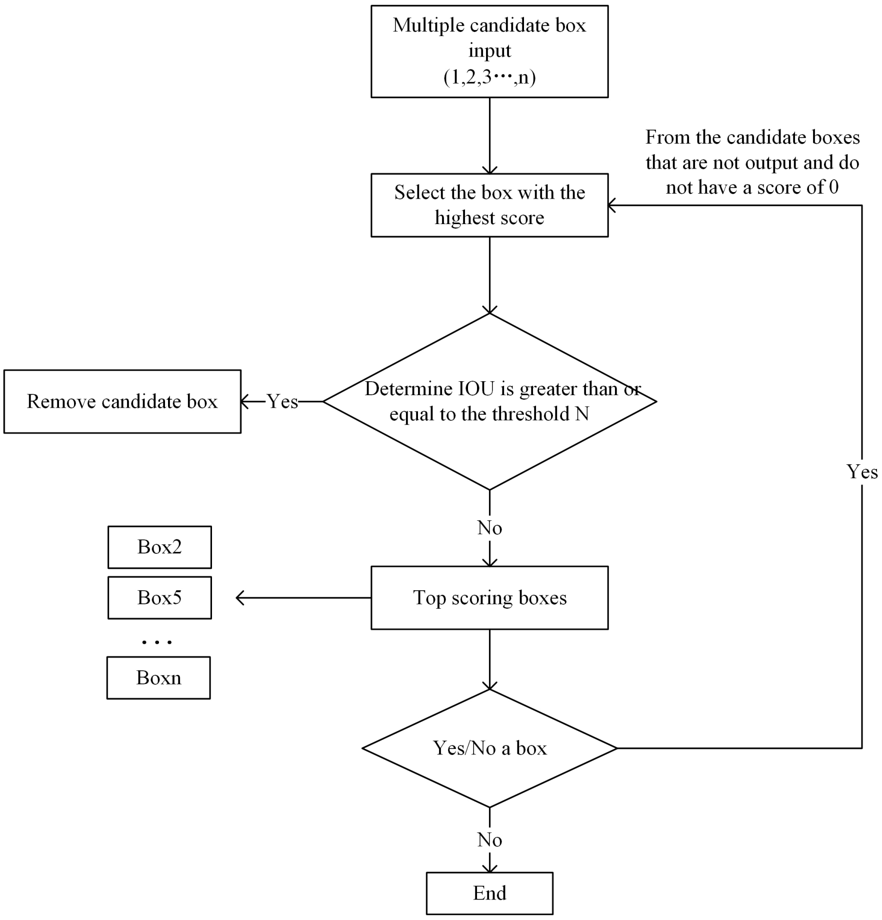

Multiple candidate box input (1,2,3⋯,n)
Select the box with the highest score
From the candidate boxes that are not output and do not have a score of 0
Determine IOU is greater than or equal to the threshold N
Remove candidate box
Yes
Yes
No
Top scoring boxes
Box2
Box5
. . .
Boxn
Yes/No a box
No
End

Figure 8. Structure of NMS.

In NMS, when the IoU value of the remaining detection frame and the maximum score detection frame are greater than the set threshold, their scores are forced to zero, which easily causes the real target to be undetected when it appears in the overlapping region. For this reason, the NMS algorithm of YOLOv5 is improved by using a flexible detection frame algorithm, Soft-NMS. Using the score and overlap information of the original candidate frames with lower scores, the new candidate frame scores are obtained by a weighted form of downscaling so that the two adjacent detection frames remain in the sequence of target detection instead of being directly set to 0 and then rejected, and the higher the overlap area of the candidate frames, the lower the score, while the score of frames with very small overlap is not affected too much. The score is calculated as in Equation (8). The formula for the Gaussian function f(x) is Equation (9), its flow chart is shown in Figure 9.

$$F(s_i) = \begin{cases} s_i, & \mathrm{IoU}(M, b_i < N) \\ s_i(1 - \mathrm{IoU}(M, b_i)), & \mathrm{IoU}(M, b_i \geqslant N) \end{cases} \tag{8}$$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{9}$$

$$s_i = s_j \times e^{-\frac{\mathrm{IoU}(M, b_i)^2}{\sigma}} \tag{10}$$

where s_i is as in Equation (10) (here, s_i is taken as the product of the $\mathrm{IoU}(M, b_i)$ Gaussian exponent and the original score s_j is taken to satisfy that the function $F(s_i)$ is close to continuous at point N). Theoretically, the smaller the threshold N, the less the overlap of candidate frames, the better the diversity of candidate frames will be. At the same time, it will produce fewer candidate regions, which leads to worse detection effect. Taking this into account, N is taken as 0.3 in this paper.

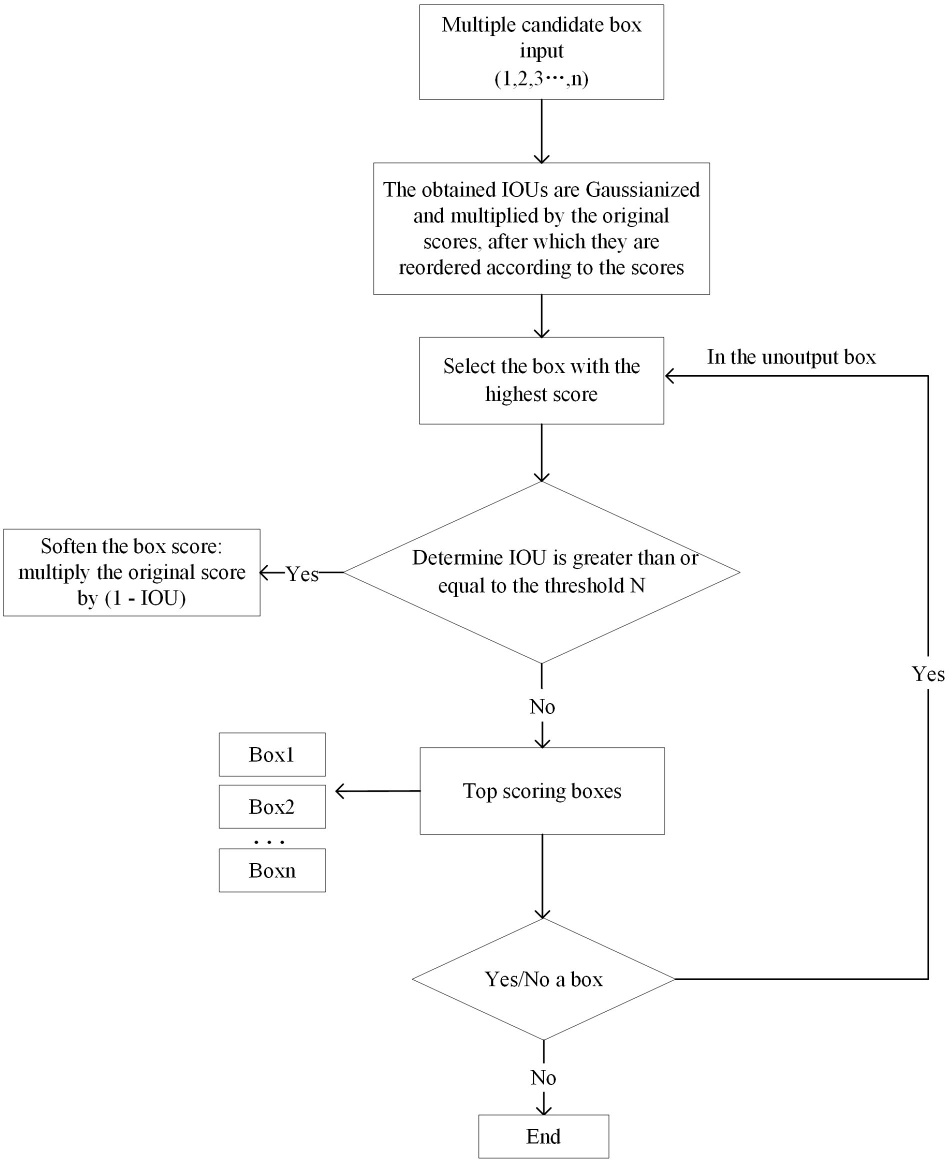

Multiple candidate box input (1,2,3⋯,n)
The obtained IOUs are Gaussianized and multiplied by the original scores, after which they are reordered according to the scores
Select the box with the highest score
In the unoutput box
Soften the box score: multiply the original score by (1 - IOU)
Yes
Determine IOU is greater than or equal to the threshold N
No
Yes
Box1
Box2
. . .
Boxn
Top scoring boxes
Yes/No a box
No
End

Figure 9. Structure of Soft-NMS.

3.4. Improved YOLOv5 Algorithm Structure

In summary, the overall structure of the improved YOLOv5 algorithm is shown in Figure 10, and the algorithm flow is shown in Figure 11. In the improved network, the input image is first cropped to 640x640 size without distortion, and then fed into the enhanced backbone feature extraction network to output four feature layers with sizes of 80, 40, 20, and 10, respectively, followed by feeding the four features into the dual-scale feature fusion network to better fuse the multi-scale features and improve the detection of large and small targets. Finally, the decoding of the original image and the feature map is used to locate the insulator and its broken area.

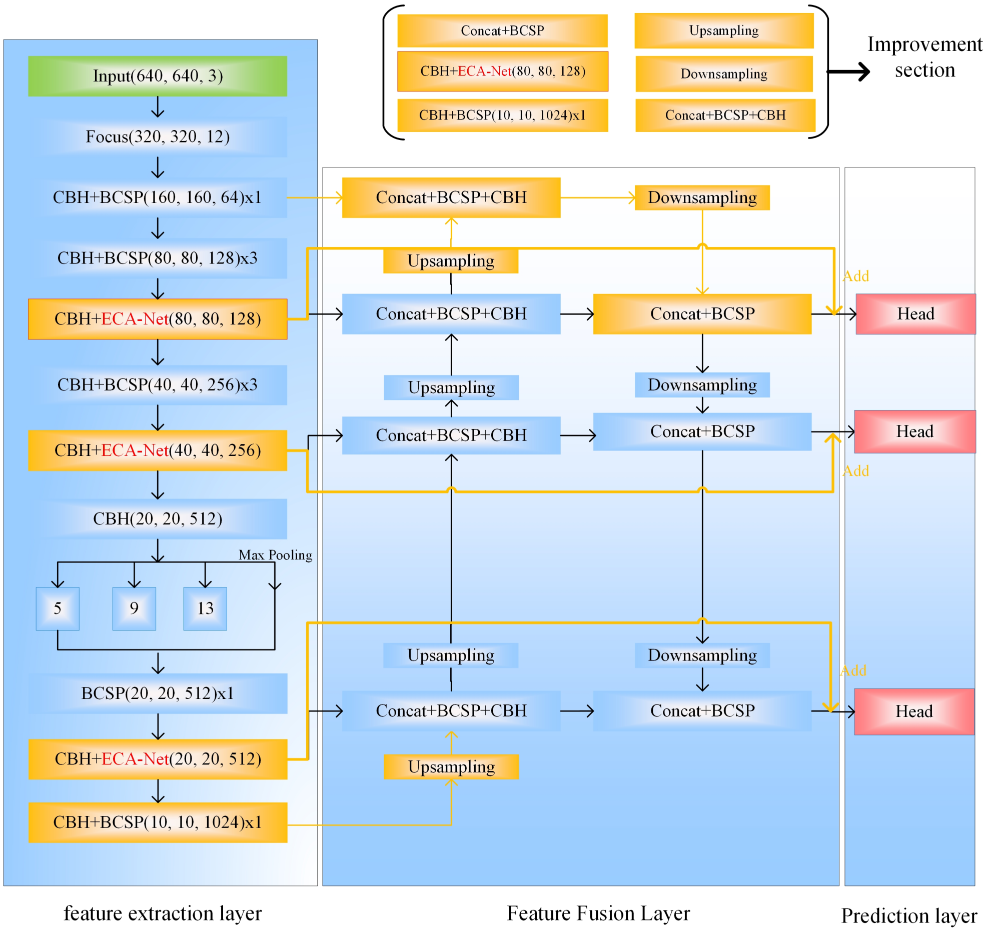


Figure 10. Improved YOLOv5 network diagram.

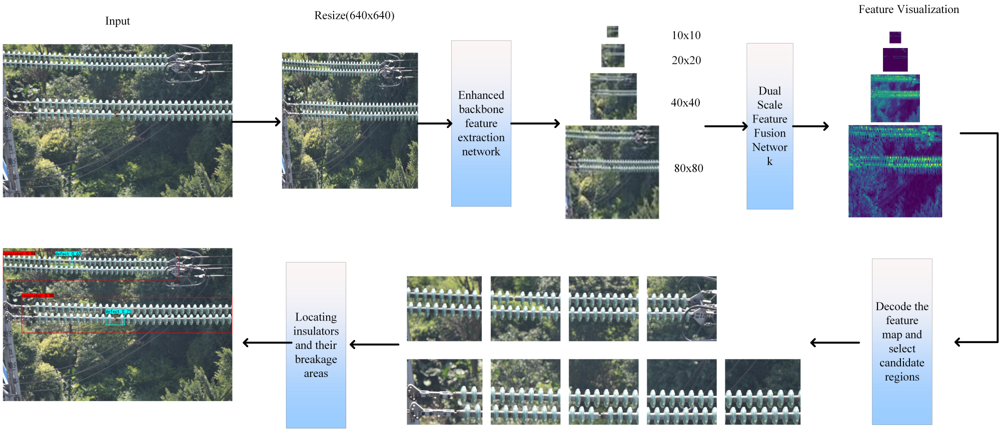
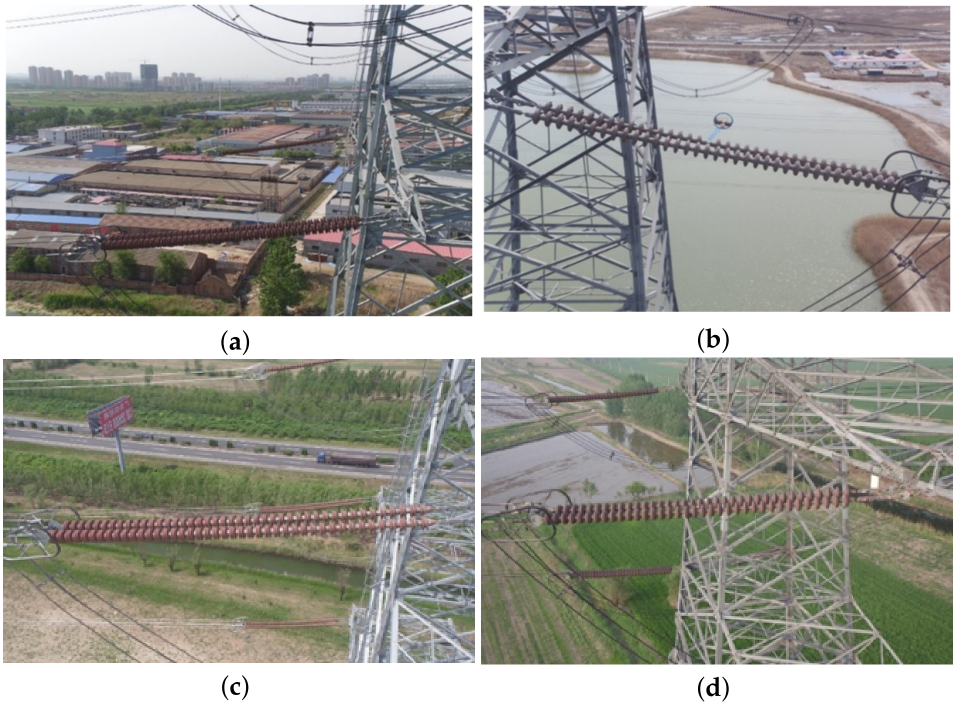

Input
Resize(640x640)
Enhanced backbone feature extraction network
10x10
20x20
40x40
80x80
Dual Scale Feature Fusion Networ k
Feature Visualization
Decode the feature map and select candidate regions
Locating insulators and their breakage areas

Figure 11. Improve network calculation process.

4. Example Analysis

4.1. Insulator Aerial Image Data Processing

The dataset used by the algorithm has 1800 images, partly from the public dataset [48] and partly from the field collection, including 800 images of broken insulator targets and 1000 images of normal insulators. The ratio of the training set, validation set, and test set is 8:1:1. Figure 12 shows several representative aerial UAVs images, (a) with insulator-like house background, (b) with mainly broken insulators and small targets with insulator-like tree background, (c) and (d) with inconsistent target size and target overlap due to different shooting angles and different insulator distances and proximity.



(**a**) (**b**) (**c**) (**d**)

Figure 12. (**a**) Insulators similar to the background of the house. (**b**) Damaged insulator; (**c**) multi-scale insulator target; (**d**) overlapping insulator.

4.2. Experimental Environment

Deep learning framework based on PyTorch 1.6 environment, Ubuntu 16.08 system, Python as 3.6.8, CUDA = 10.0, where the training graphics card configuration is 2 NVIDIA TITAN RTX 24G memory graphics cards. The local computer NVIDIA GeForce RTX 2060 SUPER 8G was used for the trained model test.

4.3. Experimental Procedure

Before training, the dataset needs to be clustered by the K-means algorithm. Compared with the a priori frame of pre-trained weights, the clustering a priori frame for this data can be more consistent with the target distribution of this data, so as to improve the detection accuracy of the model, and the clustering results are shown in Figure 13.

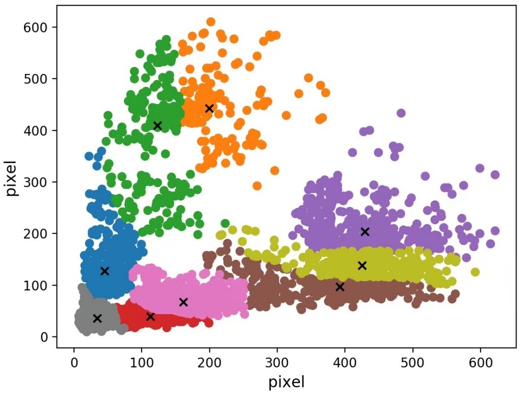
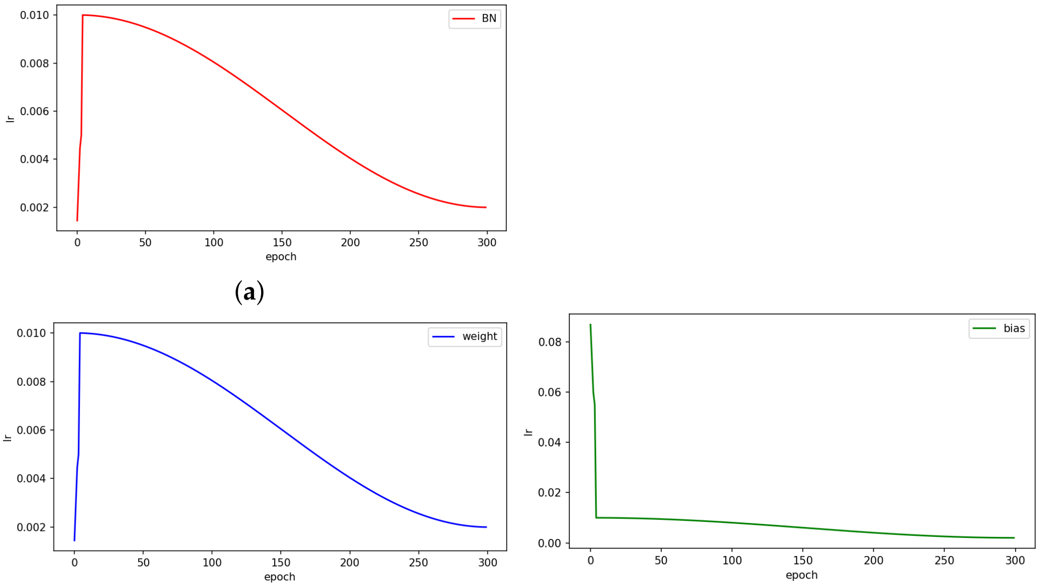


Figure 13. K-means clustering prior frame distribution.

For the YOLO series base algorithm, the migration learning idea is used to initialize its backbone part by the pre-training model of YOLOv5 obtained on the COCO dataset and loaded using the pre-training model for freeze training and thaw training. The freeze layer is trained with 50 epoch rounds and a batch-size of 16, and the thaw training with 300 rounds and a batch-size of 8. For the improved algorithm, the model is retrained with the same setting of 300 rounds. For the learning rate setting, the model uses a cosine annealing learning rate for the weight layer and BN layer, respectively, and linear decay learning rate for the bias layer. Furthermore, Warmup is used in the update of the learning rate to go for the warm-up of the learning rate, and the process first uses 1/10th of the size of the preset initial learning rate for the initial training, and then adjusts to the initial set learning rate after five generations of training. It is mainly used for initial training when the weights of the model are initialized randomly, and at this time, if a larger learning rate is chosen, it may bring about instability of the model, i.e., oscillation of the model. As for the model bias, since it has less impact on the model, a larger learning rate is used in the warm-up layer to make the parameters update faster so as to improve the overall convergence speed of the model. The learning rate curve is shown in Figure 14 below.

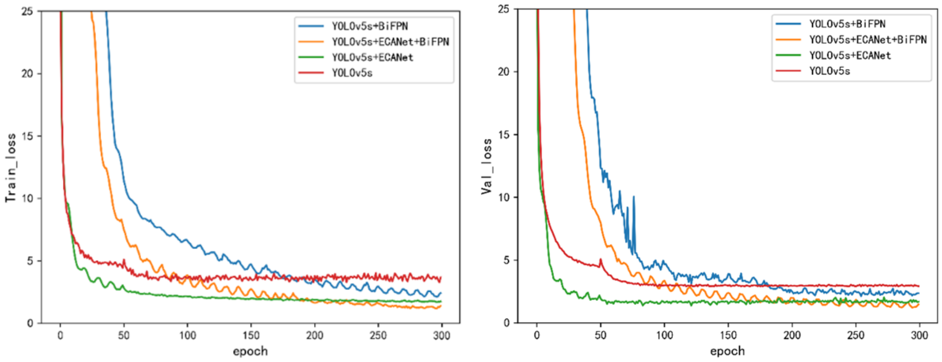


(**a**)

(**b**) (**c**)

Figure 14. (**a**) Weighted BN layer learning rate curve. (**b**) Learning rate curve of the weighting layer. (**c**) Learning rate curve of the bias layer.

From Figure 15, it can be seen that YOLOv5s converges the fastest, and the Ours 1 (YOLOv5s+ECA-Net+Soft-NMS) also converges faster and basically leveled off at 100 rounds. It can be seen that the Ours 2 (YOLOv5s+Bi-FPN+Soft-NMS) and the Ours 3 fusion algorithm (YOLOv5s+ECA-Net+Bi-FPN+Soft-NMS) converge slower and level off at about 200 rounds, which is due to the increase in computation due to the increase in model size, and it sacrifices the convergence speed in exchange for higher accuracy.



(**a**) (**b**)

Figure 15. (**a**) Training loss convergence for each of the improved models. (**b**) Validation loss convergence for each of the improved models.

4.4. Experimental Results

In this paper, Precision (PR), Recall (RE), Average Precision (AP), and mAP are used as the performance metrics of the algorithm.

(1) Precision (PR):

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{11}$$

As in Equation (11), TP denotes positive samples that are correctly classified, and FP denotes negative samples that are incorrectly classified. The specific meaning of Precision is the proportion of the part of the classifier that is considered positive and is indeed positive to the proportion of the classifier that is considered positive.

(2) Recall (RE):

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{12}$$

As in Equation (12), FN denotes the positive samples that are misclassified. Recall means specifically the fraction of the classifier that is considered positive and is indeed a positive class as a proportion of all that is indeed a positive class. AP is the sum of the area of the curve enclosed by Precision and Recall. It reflects the average accuracy of the model in detecting the target. mAP is the average of the AP of the target to be classified.

Table 2 compares the metrics of YOLOv3, YOLOv4, YOLOv5s, and their improved algorithms in insulator detection and their breakage identification, with the following conclusions:

(1) YOLOv5s has faster detection speed compared to YOLOv3 and YOLOv4. Overall, YOLOv5s has better mAP than the YOLOv3 algorithm, but slightly lower than YOLOv4 by 0.48%. However, the accuracy of YOLOv5s in insulator target detection is comparable to that of YOLOv4, both about 93.3%, but is inferior to YOLOv4 in insulator breakage (small target) detection, and its advantages are mainly reflected in its lower algorithm complexity and faster detection speed, and its lightweight feature makes it more suitable for embedded applications.

(2) Ours 1 (YOLOv5s+ECA-Net+Soft-NMS) has an overall mAP of 94.93%, an improvement of nearly 3% over the original network. Mainly for images with complex backgrounds, the accuracy rates of both insulators and their breakage are greatly improved with only 0.3 MB increase in algorithm complexity. Compared with YOLOv4, the AP of insulator detection is improved by nearly 1% to 94.35%, and the AP of breakage recognition is improved by nearly 4% to 95.51%, indicating that the improved algorithm 1, after mitigating the effect of complex backgrounds of images, also helps to improve the performance of the algorithm for small target recognition.

(3) Ours 2 (YOLOv5s+BiFPN+Soft-NMS) mAP improved to 92.88%. The improvement is mainly for the small target of insulator breakage, and the detection accuracy of broken insulators is improved from 90.63% to 93.19% compared to YOLOv4, which has a significant effect. The overall performance of the improved algorithm 2 is still improved despite the slight decrease in AP for insulator detection.

(4) Ours 3 (YOLOv5s+ECA-Net+BiFPN+Soft-NMS) superimposed the fusion of Ours 1 and 2, and its mAP improved from 91.96% to 95.02%, outperforming Ours 1 by nearly 0.1% and outperforming Ours 2 by nearly 2.14%. Compared to the original YOLOv5s, the fusion algorithm improves the accuracy by 3.06% by sacrificing about 30% of the detection speed. Moreover, the fused model is 10% smaller than the YOLOv5m model, and the mAP is 1% higher.

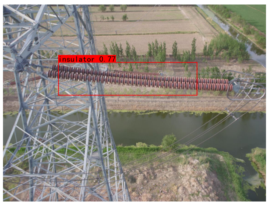
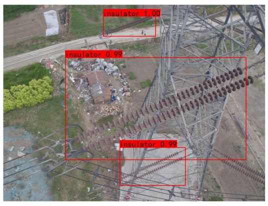
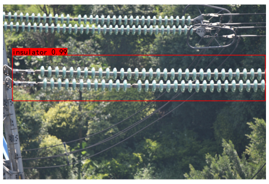
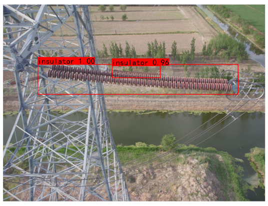
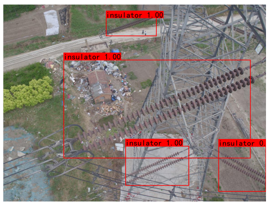
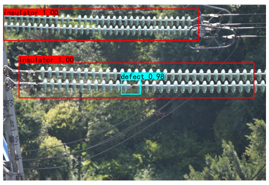
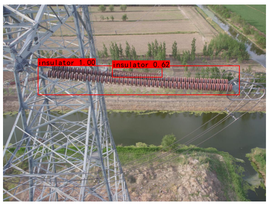
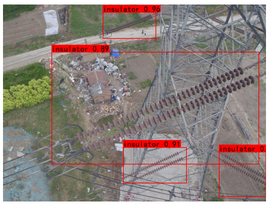
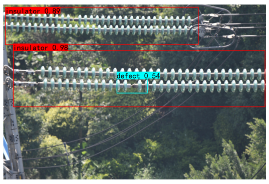
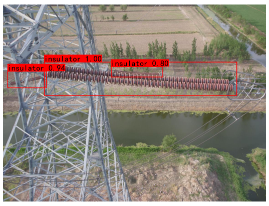
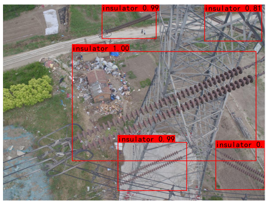
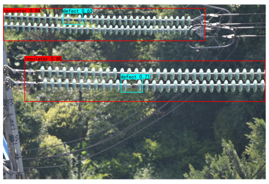
The detection results are shown through the images of the test set, and, as shown in Table 3, image 1 is a typical overlapping target image. The fusion algorithm, which is able to detect smaller targets that are obscured by the pole tower, has the best performance. Image 2 is a typical multi-target image, and the fusion algorithm is able to detect all targets. Image 3 is a typical multi-scale target image; large and small targets exist at the same time, and the fusion algorithm can be applicable to the detection of both large and small targets, which verifies the effectiveness of this model improvement.

Table 2. Comparison of different algorithms (with a confidence threshold of 0.5).

Models	Insulators (AP%)	Insulator Broken (AP%)	(mAP%)	Model Size (MB)	Video Detection Speed (FPS)
YOLOv3	93.21	88.34	90.78	235	25.00
YOLOv4	93.32	91.56	92.44	244	22.21
YOLOv5s	93.30	90.63	91.96	**27.8**	**68.18**
YOLOv5m	93.52	94.50	94.01	83.2	40.62
Ours1	94.35	**95.51**	94.93	28.1	63.81
Ours2	92.57	93.19	92.88	74.6	53.02
Ours3	**94.68**	95.36	**95.02**	74.8	49.40

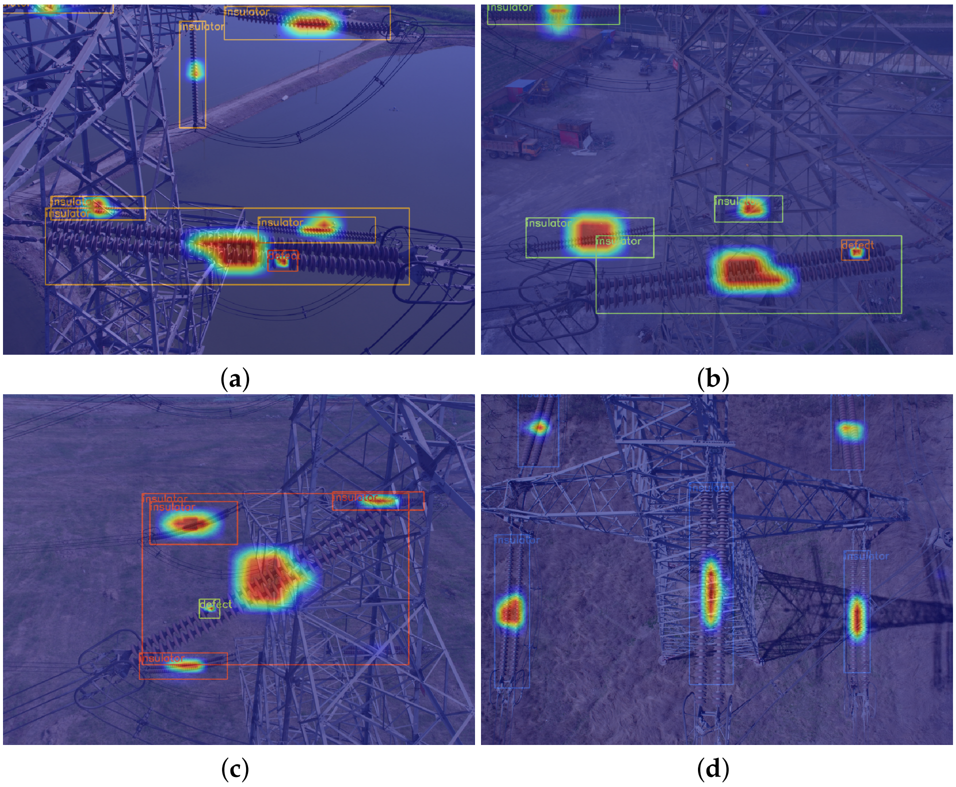
The confidence threshold is taken as 0.5 in the table. In this paper, insulator and insulator breakage are two targets, and when the probability of detecting this target is greater than 0.5, it means that the probability of the other one is lower than 0.5, so the category with probability greater than 0.5 is taken as the final predicted category. The threshold value here is not the same as the threshold value taken in Soft NMS. Furthermore, the improved models all add the Soft NMS algorithm.

Table 3. Test set image detection results.

Models	Image 1	Image 2	Image 3
YOLOv5s			
Ours 1			
Ours 2			
Ours 3			

To further visualize the regions of interest in the model, the paper concludes with a score heat map visualization analysis of the predicted results of the fusion model [49].

As can be seen from Figure 16, the heat map focuses on the center of the target detection frame, with the red color being the central focus area, and the proportion of attention spreading outward decreases. It shows that the area of concern of the model is exactly the area to be detected, which further verifies the effectiveness of the model improvement.



(**a**) (**b**)

(**c**) (**d**)

Figure 16. (**a**–**d**) Predicted thermal diagram under various insulators.

5. Conclusions

In this paper, based on YOLOv5, an enhanced channel feature extraction mechanism is embedded in its backbone feature extraction network to improve the multi-target detection capability in complex backgrounds. In the feature fusion layer, a bidirectional feature fusion network is used to enhance the recognition capability of small targets with broken insulators. In the prediction layer, the traditional candidate frame selection algorithm NMS is replaced with a more flexible Soft-NMS to improve it. In the prediction layer, the traditional candidate frame selection algorithm NMS is replaced with a more flexible Soft NMS to improve the recognition of overlapping insulator targets. The experimental results show that the recognition accuracy of the improved algorithm increases from 91.96% to 95.02%, and the detection frame rate reaches 49.4 frames/s, which confirms the feasibility and effectiveness of this paper.

The algorithm proposed in this paper is currently used for the detection of insulator self-burst and drop string problems due to high voltage, and will be followed by research on overcurrent, and flicker-induced insulator flashover burn cracks. Finally, the algorithm is expected to be extended for effective monitoring of other major power equipment and devices, and to achieve cloud-side collaboration, placing part of the target detection function at the lightweight side end.

Author Contributions: Conceptualization, G.H. and M.H.; methodology, G.H.; software, M.G.; validation, M.H., M.G. and J.Y.; formal analysis, G.H.; writing—original draft preparation, G.H. and M.G.; writing—review and editing, M.H.; visualization, M.G.; supervision, K.L. and L.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key R & D Program of China (No. 2020YFB0905900).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Dataset link: https://github.com/InsulatorData/InsulatorDataSet, accessed on 11 April 2022.

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

References

1. Bo, T.; Qiao, Q.; Li, H. Aerial image recognition of transmission line insulator strings based on color model and texture features. *J. Electr. Power Sci. Technol.* **2020**, *35*, 13–19.
2. Tan, P.; Li, X.F.; Xu, J.M.; Ma, J.E.; Ning, Y. Catenary insulator defect detection based on contour features and gray similarity matching. *J. Zhejiang Univ. Sci. Appl. Phys. Eng.* **2020**, *21*, 64–73. [CrossRef]
3. Wu, Q.; An, J. An Active Contour Model Based on Texture Distribution for Extracting Inhomogeneous Insulators From Aerial Images. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 3613–3626. [CrossRef]
4. Wang, H.; Cheng, L.; Liao, R.; Zhang, S.; Yang, L. Nonlinear Mechanical Model of Composite Insulator Interface and Nondestructive Testing Method for Weak Bonding Defects. *Zhongguo Dianji Gongcheng Xuebao/Proceedings Chin. Soc. Electr. Eng.* **2019**, *39*, 895–905.
5. Tian, Y.; Si, Y.; Mengyu, X.; Lisong, Y.; Zhengyan, L.; Xudong, X. Laser detection method for cracks in glass insulators. *Power Grid Technol.* **2020**, *44*, 3156–3163.
6. Zijian, Z.; Enji, M.; Xufeng, L.; Youtong, F. Insulator fault detection based on deep learning with Hu-invariant moments. *J. Railw.* **2021**, *43*, 71–77.
7. Zhai, Y.; Chen, R.; Yang, Q.; Li, X.; Zhao, Z. Insulator Fault Detection Based on Spatial Morphological Features of Aerial Images. *IEEE Access* **2018**, *6*, 35316–35326. [CrossRef]
8. Liu, T. Porcelain Insulator Crack Location and Surface States Pattern Recognition Based on Hyperspectral Technology. *Entropy* **2021**, *23*, 486.
9. Rahman, E.U.; Zhang, Y.; Ahmad, S.; Ahmad, H.I.; Jobaer, S. Autonomous Vision-Based Primary Distribution Systems Porcelain Insulators Inspection Using UAVs. *Sensors* **2021**, *21*, 974. [CrossRef]
10. Siddiqui, Z.A.; Park, U. A Drone Based Transmission Line Components Inspection System with Deep Learning Technique. *Energies* **2020**, *13*, 3348. [CrossRef]
11. Choi, I.H.; Koo, J.B.; Son, J.A.; Yi, J.S.; Yoon, Y.G.; Oh, T.K. Development of equipment and application of machine learning techniques using frequency response data for cap damage detection of porcelain insulators. *Appl. Sci.* **2020**, *10*, 2820. [CrossRef]
12. Zhai, Y.; Cheng, H.; Chen, R.; Yang, Q.; Li, X. Multi-saliency aggregation-based approach for insulator flashover fault detection using aerial images. *Energies* **2018**, *11*, 340. [CrossRef]
13. Hosseini, M.M.; Umunnakwe, A.; Parvania, M.; Tasdizen, T. Intelligent Damage Classification and Estimation in Power Distribution Poles Using Unmanned Aerial Vehicles and Convolutional Neural Networks. *IEEE Trans. Smart Grid* **2020**, *11*, 3325–3333. [CrossRef]
14. Davari, N.; Akbarizadeh, G.; Mashhour, E. Intelligent Diagnosis of Incipient Fault in Power Distribution Lines based on Corona Detection in UV-Visible Videos. *IEEE Trans. Power Deliv.* **2020**, *36*, 3640–3648. [CrossRef]
15. Wei, C.; Peiwei, X.; Zhiyong, Y.; Xinhao, J.; Bo, J. Detection of small and dim targets in infrared images under complex background. *Appl. Opt.* **2021**, *42*, 643–650. [CrossRef]
16. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
17. 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]
18. Akbari, M.; Liang, J. Semi-recurrent CNN-based VAE-GAN for sequential data generation. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 2321–2325.
19. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
20. Jianpeng, B.; Fan, L.; Peixu, H.; Yamin, L.; Xiaoyun, S. Breakage identification and location of transmission line insulators under complex environment. *High Volt. Technol.* **2022**, *48*, 8.
21. Redmon, J.; Farhadi, A. *YOLO9000: Better, Faster, Stronger*; IEEE: Piscataway, NJ, USA, 2017; pp. 6517–6525.
22. Sabour, S.; Frosst, N.; Hinton, G.E. Dynamic routing between Capsules. Available online: https://proceedings.neurips.cc/paper/2017/hash/2cad8fa47bbef282badbb8de5374b894-Abstract.html (accessed on 11 April 2022).
23. Hai, L.; Yang, L.; Zhengrong, Z. Luminous remote sensing building area detection with complex background. *J. Infrared Millim. Waves* **2021**, *40*, 369–380.
24. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
25. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 21–37.
26. Xiaojun, L.; Wei, X.; Yunpeng, L. Small Target Detection Algorithm for UAV Aerial Photography Images Based on Enhanced Underlying Features. *Comput. Appl. Res.* **2021**, *38*, 1567–1571.
27. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.

28. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]
29. Xu, L.F.; Huang, H.F.; Ding, W.l.; Fan, Y.l. Detection of small fruit target based on improved DenseNet. *J. Zhejiang Univ. (Eng. Sci.)* **2021**, *55*, 377–385.
30. Iandola, F.; Moskewicz, M.; Karayev, S.; Girshick, R.; Darrell, T.; Keutzer, K. Densenet: Implementing efficient convnet descriptor pyramids. *arXiv* **2014**, arXiv:1404.1869.
31. Shu, Z.; Haotian, W.; Xiaochong, D.; Yurong, L.; Ye, L.; Yinxin, W.; Yingyun, S. Bolt detection technology for transmission lines based on deep learning. *Power Grid Technol.* **2021**, *45*, 2821–2828.
32. Ruisheng, L.; Yanlong, Z.; Denghui, Z.; Dan, X. Detection of pin defects in transmission lines based on improved SSD. *High Volt. Technol.* **2021**, *47*, 3795–3802.
33. Haibin, W.; Xiying, W.; Meihong, L.; Aili, W.; He, L. Combining cavity convolution and migration learning to improve the detection of dangerous goods in X-ray security inspection of yolov4. *China Opt.* **2021**, *14*, 9.
34. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [CrossRef]
35. Zhe, D.; Wenzong, L.; Fenting, Y. Research on target detection method based on multi-feature information fusion. *Comput. Appl. Softw.* **2020**, *37*, 122–126.
36. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 886–893.
37. Neubeck, A.; Van Gool, L. Efficient non-maximum suppression. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, China, 20–25 June 2006; Volume 3, pp. 850–855.
38. Tao, X.; Zhang, D.; Wang, Z.; Liu, X.; Zhang, H.; Xu, D. Detection of Power Line Insulator Defects Using Aerial Images Analyzed With Convolutional Neural Networks. *IEEE Trans. Syst. Man, Cybern. Syst.* **2020**, *50*, 1486–1498. [CrossRef]
39. Cai, Z.; Vasconcelos, N. Cascade r-cnn: Delving into high quality object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6154–6162.
40. Cristi, F. Pre-commit-ci[bot] and Glenn-jocher. Ultralytics, Yolov5. 2020. Available online: https://github.com/ultralytics/yolov5/ (accessed on 11 April 2022).
41. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 658–666.
42. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020.
43. Bodla, N.; Singh, B.; Chellappa, R.; Davis, L.S. Soft-NMS–improving object detection with one line of code. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 5561–5569.
44. Howard, A.; Sandler, M.; Chu, G.; Chen, L.C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for mobilenetv3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 1314–1324.
45. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
46. Roy, K.; Hasan, M.; Rupty, L.; Hossain, M.S.; Mohammed, N. Bi-FPNFAS: Bi-Directional Feature Pyramid Network for Pixel-Wise Face Anti-Spoofing by Leveraging Fourier Spectra. *Sensors* **2021**, *21*, 2799. [CrossRef] [PubMed]
47. Tan, M.; Pang, R.; Le, Q.V. Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 10781–10790.
48. Wang, Z. Insulator Data Set, Chinese Power Line Insulator Dataset. 2018. Available online: https://github.com/InsulatorData/InsulatorDataSet (accessed on 11 April 2022)
49. Wang, H.; Wang, Z.; Du, M.; Yang, F.; Zhang, Z.; Ding, S.; Mardziel, P.; Hu, X. Score-CAM: Score-weighted visual explanations for convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 16–17 June 2020; pp. 24–25.