Hindawi
Scientific Programming
Volume 2021, Article ID 6235452, 12 pages
https://doi.org/10.1155/2021/6235452

Research Article

Multitarget Detection of Transmission Lines Based on DANet and YOLOv4

Zhen Yang[1], **Xuefei Xu**[1], **Keke Wang**[2], **Xin Li**[1] and **Chi Ma**[1]

[1]*Faculty of Electrical and Control Engineering, Liaoning Technical University, Huludao 125105, China*
[2]*Nanjing Institute of Electronic Technology, Nanjing 201139, China*

Correspondence should be addressed to Xuefei Xu; 1797511838@qq.com

Received 8 July 2021; Revised 24 November 2021; Accepted 6 December 2021; Published 17 December 2021

Academic Editor: Jiangbo Qian

In order to accurately identify targets such as insulators, shock hammers, bird nests, and spacers on high-voltage transmission lines, this paper proposes a multitarget detection model for transmission lines based on DANet and YOLOv4. First, the DANet and YOLOv4 are fused to solve the difficulty in understanding the scene and the discrimination of pixels caused by the complex and diverse scenes of UAV' (unmanned aerial vehicle) aerial images (lighting, viewing angle, scale, occlusion, and so on) so as to improve the significance of the detection target. Gaussian function and KL (Kullback–Leibler) divergence are used to improve the nonmaximum suppression in YOLOv4 so as to improve the recognition rate of occluded targets; the focal loss function and the balanced cross entropy function are used to improve the loss function of YOLOv4 in order to reduce the impact of not only the imbalance between the background and the detection target but also the imbalance among the samples, which is aimed at improving the accuracy of the detection. Then, a data set is made for the experiment by using the UAV inspection image provided by a power grid company in Eastern Inner Mongolia. Finally, the algorithm proposed in this paper is compared with other target detection algorithms. Experimental results show that the average detection accuracy of the proposed algorithm can reach 94.7%, and the detection time of each image is 0.05 seconds. The method has good accuracy, real-time, and robustness.

1. Introduction

The insulators, shock hammers, and spacers are all important components on high-voltage transmission lines, which are vital to the stable operation of high-voltage transmission lines [1]. Because high-voltage transmission lines adopt overhead transmission, insulators and other important power components are exposed to the natural environment for a long time, so they are easily affected by the weather and natural environment. In order to ensure the stable operation of the power system, it is necessary to carry out regular inspection of transmission lines, insulators, and other electrical equipment for timely maintenance. Currently, the use of unmanned aerial vehicles for transmission line inspection has become an important way which mainly uses drones to take a large number of pictures and relies on professional inspectors to inspect these pictures. Due to the use of artificial picture identification, with the increase in the number of pictures, this approach will not only spend a lot of manpower and resources but will also cause misses and unable to detect existence of faults. With the rapid development of deep learning in the field of image recognition, it has become a trend and research hotspot to apply image recognition technology to power inspection [2].

Target detection algorithms based on deep learning can be divided into two categories according to different detection stages. The first type is the two-stage algorithm. This type of algorithm needs to generate the region proposal of the candidate region that may contain the object and then further classify and calibrate the candidate region to obtain the final detection result. The representative algorithms mainly include RCNN [3], Fast RCNN [4], Faster RCNN [5], and RFCN [6]. The recognition accuracy of this algorithm is high, but the training time of the model is long, and the detection speed is slow. For transmission line inspection, it is necessary to find and eliminate hidden dangers in time to

ensure the stable operation of the power system, so the detection speed must be improved. The second type is the one-stage algorithm. This type of method directly trains the network end-to-end and directly gives the final detection result without going through the steps of candidate regions. The representative algorithms mainly include SSD [7], YOLOv1 [8], YOLOv2 [9], and YOLOv3 [10]. This kind of algorithm is characterized by fast detection speed and little interference of background to target recognition but low accuracy of recognition, especially for small target recognition. Power line inspection must be able to accurately identify all kinds of power equipment. This kind of algorithm has low recognition accuracy for small power equipment such as screws and pins, so it needs to be further optimized. Zhao [11] et al. proposed a method called AVSCNet for pin-missing in transmission lines, which only optimized the network based on the structural characteristics of bolts in a certain background, without verifying the detection effect in different shooting angles and different natural environments. Tao [12] et al. proposed a novel deep CNN cascading architecture for performing localization and detecting defects in insulators, but the speed of this method is relatively slow. Liu [13] et al. used YOLOv3 to identify insulators with an accuracy of only 88.7%. Wang [14] et al. proposed an insulator detection algorithm based on Gaussian YOLOv3. Experimental results show that the improved YOLOv3 algorithm can accurately locate the position of the object. The detection accuracy of insulators in the test set reaches 93.8%.

In the actual transmission line inspection, the background of drone aerial images is complex, the light intensity is different, and some targets are blocked. Therefore, there are certain difficulties in accurately identifying the targets of the transmission line, which need to be resolved.

Based on the above analysis, this paper integrates DANet (dual attention network) for scene segmentation [15] and YOLOv4 [16] to detect insulators, spacers, and others in transmission lines. This article mainly improves YOLOv4 according to the following three aspects. (1) DANet is fused in the YOLOv4 algorithm. It can improve the recognition accuracy of targets in complex backgrounds and reduce the impact of different light intensities and camera angles on recognition accuracy. (2) The Gaussian function and KL divergence are used to improve nonmaximal suppression of YOLOv4 as well as improve the detection rate of the target occlusion. (3) The focal loss [17] function and the balanced cross entropy function are used to improve the loss function of YOLOv4. It can reduce the impact of not only the imbalance between the background and the detection target but also the imbalance of the samples, which will improve the accuracy of the detection.

2. Algorithm Principle

2.1. The Main Structure of YOLOv4. Compared with YOLOv3, YOLOv4 has many improvements, which are mainly reflected in the input terminal, backbone network, and the neck part of the network. In terms of input, YOLOv4 uses Mosaic data enhancement, CmBN (Cross min-Batch Normalization), and SAT (self-antagonism training). Mosaic data enhanced random zoom by using 4 pictures, then randomly distributed, and greatly enriched the detection data set. The random scaling added a lot of small goals so that the robustness of the network is better. When mosaic enhances training, the data of 4 images can be directly calculated, so the minibatch size need not to be large, a GPU can achieve a better effect, and it is reduced greatly. YOLOv4's backbone feature extraction network uses CSPDarkNet53, which is improved on the basis of YOLOv3 backbone network DarkNet53, which contains 5 CSP modules. The volume nuclear size in front of each CSP module is 3×3, and the step size stride equals to 2; therefore, it can play the role of down sampling. The use of CSPDarkNet53 enhances the learning ability of the model so that the accuracy is maintained while lightweight, and the main network structure of the CSPDarkNet53 is shown in Figure 1. The neck structure of YOLOv4 mainly adopts the SPP (spatial pyramid pooling) [18] module and the way of FPNs (feature pyramid networks) [19] + PAN (path aggregation network). The SPP module uses a maximum pooling mode of $k = \{1 \times 1, 5 \times 5, 9 \times 9, 13 \times 13\}$ and then performs the concatenate operation of different scales. This approach increases the receiving range of the backbone characteristic than a simple method using $k \times k$ largest pooling. The most important context feature is remarkably separated. YOLOv4 also adds a feature pyramid from the bottom to the top after the FPN. FPN layer conveys strong semantic features from top to bottom, while feature pyramid conveys strong location features from bottom to top. By combining the two of above, parameters of different detection layers can be aggregated from different backbone layers, so as to extract features better.

In order to solve the impact of complex and diverse scenes (illumination, perspective, scale, occlusion, and so on) of UAV aerial images on detection targets, this article integrates the dual attention network on the basis of YOLOv4 network, and the Gaussian function is used to improve the nonmaximum suppression. It uses the focal loss function and the balanced cross entropy function to improve the loss function of the network so that the recognition accuracy can be improved. The detection flow chart of the algorithm of this paper is shown in Figure 2.

The first step is to label the aerial image taken during the drone inspection with labelImg and then adjust the input size of the image to 608×608. The second step is to input the processed pictures into the improved YOLOv4 network for training and perform multiple rounds of training to obtain the training weights of the transmission line insulator detection model. The final step is using the test set to verify this model.

2.2. The Structure of Dual Attention Networks. The overall framework of DANet is shown in Figure 3. ResNet is deformed; that is, after the down sampling of the last two modules is removed, an output feature map with the size of 1/8 of the input image is obtained by using void convolution, and then the output feature map is input to the two attention

Type	Filters	Size	Output	
Convolutional	32	3*3	256*256	
Convolutional	64	3*3*/2	128*128	
Cross Stage Partial				
Convolutional	32	1*1		*1
Convolutional	64	3*3		
Residual			128*128	
Convolutional	128	3*3*/2	64*64	
Cross Stage Partial				
Convolutional	64	1*1		*2
Convolutional	64	3*3		
Residual			64*64	
Convolutional	256	3*3*/2	32*32	
Cross Stage Partial				
Convolutional	128	1*1		*8
Convolutional	128	3*3		
Residual			32*32	
Convolutional	512	3*3*/2	16*16	
Cross Stage Partial				
Convolutional	256	1*1		*8
Convolutional	256	3*3		
Residual			16*16	
Convolutional	1024	3*3*/2	8*8	
Cross Stage Partial				
Convolutional	512	1*1		*4
Convolutional	512	3*3		
Residual			8*8	
Avgpool		Global		
Connected		1000		
softmax				

Figure 1: CSPDarknet53 network structure diagram. The multi-objective detection method of transmission line is based on DANet-YOLOv4.

modules, respectively, to capture global semantic information (some relationship established between pixels). In the positional attention module, firstly, a positional attention matrix is used to model the relationship between any two points. Then, the attention matrix is multiplied by the characteristic matrix, and the multiplied results and the original feature matrix are elements' additions resulting in the final result of a certain characteristic ability to global semantics. The operation of the channel payment module is similar, but multiplication is calculated at the channel dimension. Finally, the results of the two modules are aggregated to get a better characterization result for the next pixel prediction.

2.3. Position Attention Module. The DANet is a network that applies self-attention. It introduces a self-attention mechanism to capture the feature dependency in the spatial dimension and the channel dimension, respectively. As you can see from its structure diagram, it is composed of two parallel attention modules. The first one gets the dependency relationship between any two positions in the feature map, which is called position attention module (PAM). Its structure is shown in Figure 4.

As illustrated in Figure 4, a local feature $A \in R^{C\times H\times W}$ is given; we firstly feed it into a convolution layers to generate two new feature maps B and C, respectively, where$\{B, C\} \in R^{C\times H\times W}$. Then, we reshape them to $R^{C\times N}$, where $N = H\times W$ is the number of pixels. After that we perform a matrix multiplication between the transpose of C and B and apply a softmax layer to calculate the spatial attention map $S \in R^{N\times N}$:

$$S_{j,i} = \frac{\exp\left(B_i \cdot C_j\right)}{\sum_{i=1}^{N} \exp\left(B_i \cdot C_j\right)}, \tag{1}$$

where s^{ji} represents the elements in the i^{th} column and j^{th} of S. The more similar feature representations of the two position contribute to greater correlation between them.

Meanwhile, we feed feature A into a convolution layer to generate a new feature map $D \in R^{C\times H\times W}$ and reshape it to $R^{C\times N}$. Then, we perform a matrix multiplication between D and the transpose of S and reshape the result to $R^{C\times H\times W}$. Finally, we multiply it by a scale parameter α and perform an element-wise sum operation with the features A to obtain the final output $E \in R^{C\times H\times W}$ as follows:

$$E_j = \alpha \sum_{i=1}^{N} \left(s_{ji} D_i\right) + A_j, \tag{2}$$

where D_i is the i^{th} column and α is the training parameter.

It can be inferred from Equation (2) that the result feature E_j of each position is the feature and weighted sum of all positions. Therefore, it has a global context characteristic and selectively aggregates contexts according to spatial attention graphs. The similar semantic features reinforce each other, thus promoting the semantic consistency of the same category of objects and greatly improving the characteristic information of insulators, shock hammer, and spacers affected by background interference.

2.4. Channel Attention Module. Another attention module represents the dependency between any two channels, called channel attention module (CAM), and its structure is shown in Figure 5.

Different from the position attention module, we directly calculate the channel attention map $\mathbf{X} \in \mathrm{R}^{C\times C}$ from the original features $\mathbf{A} \in R^{C\times H\times W}$. Specifically, we reshape $\mathbf{A}$ to $R^{C\times N}$ and then perform a matrix multiplication between $\mathbf{A}$ and the transpose of $\mathbf{A}$. Finally, we apply a softmax layer to obtain the channel attention map $\mathbf{X} \in \mathrm{R}^{C\times C}$:

$$x_{j,i} = \frac{\exp\left(A_i \cdot A_j\right)}{\sum_{i=1}^{C} \exp\left(A_i \cdot A_j\right)}, \tag{3}$$

where x^{ji} measures the i^{th} channel's impact on the j^{th} channel. In addition, we perform a matrix multiplication between the transpose of X and A and reshape their result to

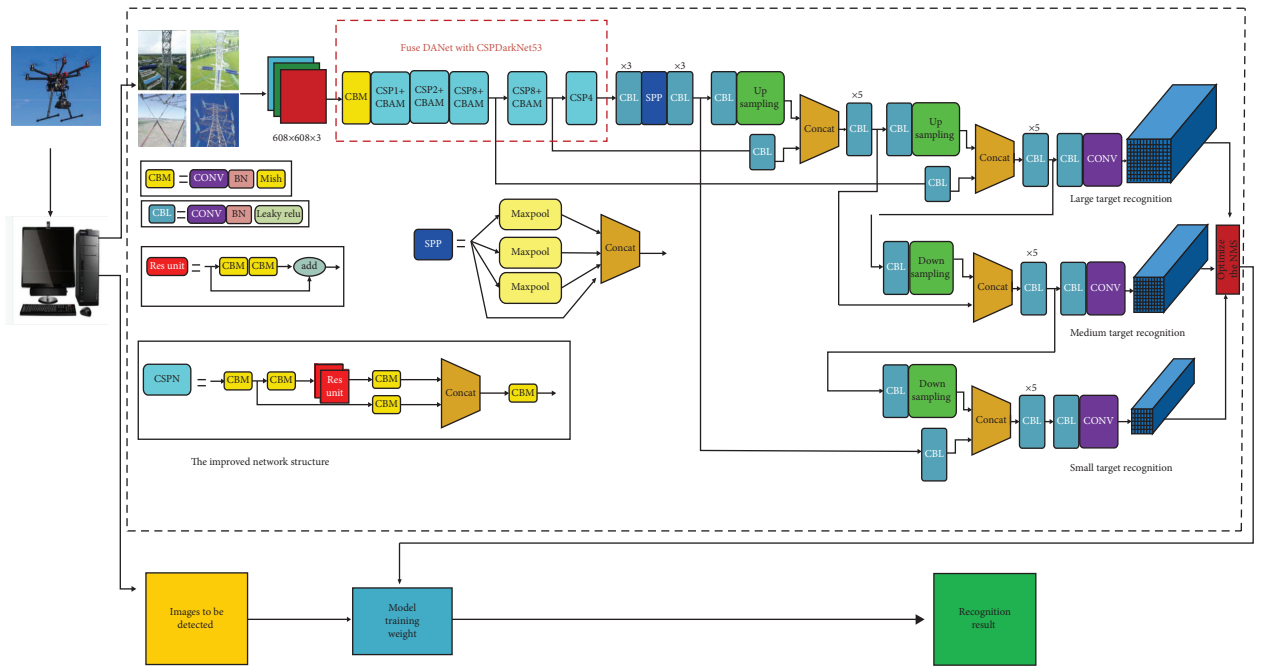

Figure 2: Detection flow chart of this article.

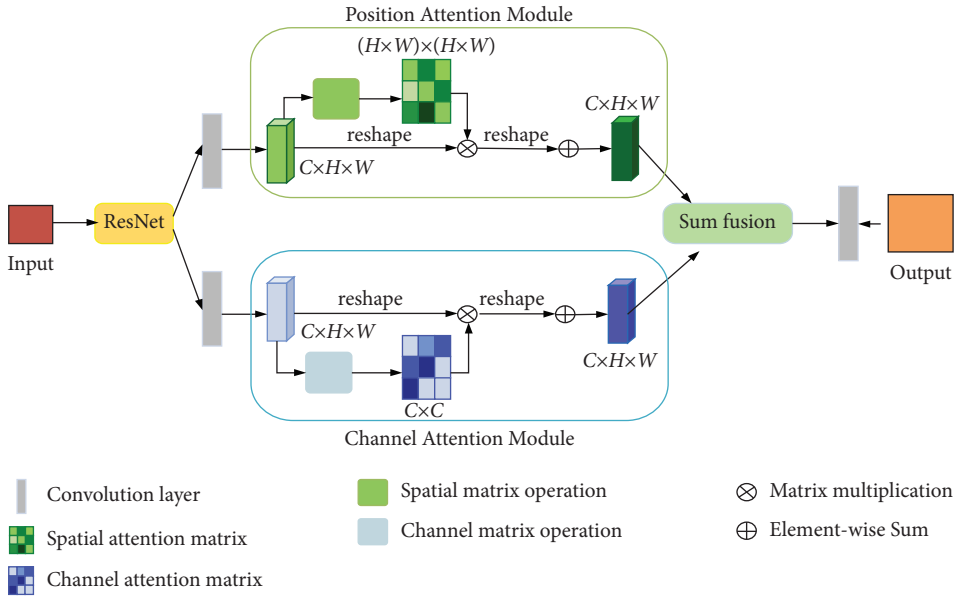

Figure 3: DANet structure diagram.

$R^{C\times H\times W}$. Then, we multiply the result by a scale parameter β and perform an element-wise sum operation with A to obtain the final output $E \in R^{C\times H\times W}$:

$$E_j = \beta \sum_{i=1}^{C} \left(x_{ji} A_i\right) + A_j, \tag{4}$$

where β is the training parameter.

Each feature channel map is corresponding to a different category. Equation (4) shows that the final feature of each channel is a weighted sum of the features of the original channel with the features of all feature channels, helping to improve the identifiability of features and thus the confidence of categories.

2.5. Detection Model Fused with DANet. In order to make the improved algorithm have higher recognition accuracy and reduce the influence of DANet on the recognition rate of YOLOv4 network, the DANet model was introduced only

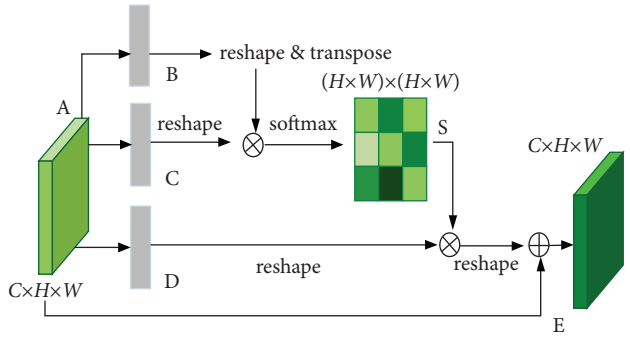

Figure 4: Position attention module.

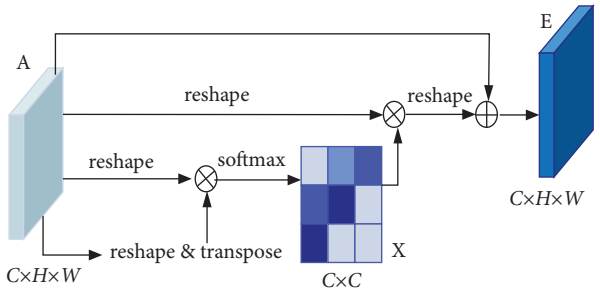

Figure 5: Channel attention module.

after the residual module in each set of convolution and residual layer of YOLOv4 network. The improved network structure is shown in Figure 6.

DANet adopts spatial attention module and channel attention module, which establish semantic interdependence of spatial dimension and channel dimension, respectively, and makes use of feature information in global view, which is crucial for scene segmentation and can more accurately distinguish targets of different categories. Therefore, the fusion of DANet in YOLOv4 can greatly improve the saliency of insulators, shock hammers, spacers, and bird nests in complex background, thus improving the accuracy of identification. The experimental data in Table 1 prove that DANet can improve the average accuracy of the model.

2.6. Improvement of Nonmaximum Suppression. Due to the complex background and camera angles of aerial image, some detected targets will be blocked. The standard NMS (nonmaximum suppression) [20] determines the fraction of the bounding box through the relation between the Intersection over Union and the size of the threshold. The calculation formula of NMS is as follows:

$$s_i = \begin{cases} s_i & \text{iou}(M, b_i) < N_t \\ 0 & \text{iou}(M, b_i) \geq N_t \end{cases}, \quad (5)$$

where S_i represents the score of each bounding box; M represents the current bounding box with the highest score; bi represents a bounding box; and N_t is the threshold value set.

Type	Filters	Size	Output	
Convolutional	32	3*3	256*256	
Convolutional	64	3*3*/2	128*128	
Cross Stage Partial				
Convolutional	32	1*1		*1
Convolutional	64	3*3		
Residual			128*128	
DANet			128*128	
Convolutional	128	3*3*/2	64*64	
Cross Stage Partial				
Convolutional	64	1*1		*2
Convolutional	64	3*3		
Residual			64*64	
DANet			64*64	
Convolutional	256	3*3*/2	32*32	
Cross Stage Partial				
Convolutional	128	1*1		*8
Convolutional	128	3*3		
Residual			32*32	
DANet			32*32	
Convolutional	512	3*3*/2	16*16	
Cross Stage Partial				
Convolutional	256	1*1		*8
Convolutional	256	3*3		
Residual			16*16	
DANet			16*16	
Convolutional	1024	3*3*/2	8*8	
Cross Stage Partial				
Convolutional	512	1*1		*4
Convolutional	512	3*3		
Residual			8*8	
DANet			8*8	
Avgpool		Global		
Connected		1000		
softmax				

Figure 6: Improved network structure.

Table 1: Comparison of test results of different models.

Algorithm	mAP (%)	Test time (sec/img)
RetinaNet	90.5	0.14
Faster RCNN	95.2	0.26
SSD	89.9	0.19
FCOS	93.8	0.09
YOLOv4	91.3	0.04
Ours	94.7	0.05

By the above formula, it can be found that when IoU (Intersection over Union) is greater than N_t, the score of the bounding box is directly 0, which is equivalent to discarding the bounding box. This will cause a missed inspection on the target of the occlusion.

The score used by NMS is only the classification confidence score, which cannot reflect the positioning accuracy of the prediction box; that is, classification confidence and localization confidence do not have consistency.

In order to solve the above problem, this paper adds a prediction of a positioning confidence based on NMS,

making the border position with high classification confidence become more accurate, thereby effectively improving the performance of the detection. First, the Gaussian function model is used to predict the coordinate position distribution of the bounding box, as shown in formula (6) [21]. Dirac delta distribution was used to model the position distribution of coordinate points of the ground truth bounding box. As shown in formula (7) [21],

$$P_{\Theta}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-x_e)^2/2\sigma^2}, \tag{6}$$

where Θ represents a collection of training parameters; x_e represents the predicted bounding box position; and σ is the standard deviation of the predicted box position.

$$P_D(x) = \delta\left(x - x_g\right), \tag{7}$$

where x_g is the location of ground truth bounding box.

The closer the distribution of the prediction box is to the distribution of the ground truth bounding box, the more accurate the target positioning of the model is. In this paper, KL (Kullback–Leibler) divergence is introduced to measure the similarity of probability distribution $P_{\Theta}(x)$ and $P_D(x)$.

In this paper, the border regression loss function (KL Loss) based on KL divergence is used to minimize the divergence between the training process and the model. The lower loss indicates that the prediction boundary distribution is closer to the real object distribution. The expression of KL loss is shown in the following equation [21]:

$$L_{\text{reg}} = e^{-\alpha}\left(\left|x_g - x_e\right| - \frac{1}{2}\right) + \frac{1}{2}\alpha, \tag{8}$$

$$\alpha = \log\left(\sigma^2\right), \tag{9}$$

where x_e represents the predicted bounding box position, x_g is the location of ground truth bounding box, and σ is the standard deviation of the predicted box position.

2.7. Improvement of the Loss Function. This paper uses the focal loss function and the balanced cross entropy function to improve the function of YOLOv4 loss. When YOLOv4 calculates the loss value, the model will divide the prediction boxes into positive samples and negative samples based on the value of CIoU. In the images taken by UAV, some insulator targets have a small size, and their proportion in the image is much smaller than that of the background, which will lead to the imbalance of positive and negative samples in the data set, as shown in the left figure in Figure 7. In the problem of insulator target detection, the insulator to be positioned is called the foreground and the other parts are called the background. There is a phenomenon of unbalanced complexity between foreground and background in the aerial insulator image sample, as shown in the middle of Figure 7. With a large number of insulators and a large number of disturbances in the background, it belongs to an indistinguishable sample with complex foreground and background. In the right of Figure 7, the number of insulators is small; the background is simple, which makes the sample easy to divide. One of the reasons for the relatively low accuracy of the one-stage algorithm is that there is a serious imbalance of the samples in the data set. Too many samples of a certain type in the data set will make it difficult for the model to learn the information of other types of samples [22].

To solve the above problems, the equilibrium cross entropy function and focal loss function were used to improve the standard cross entropy (CE) in YOLOv4 loss function. The original loss function of YOLOv4 consists of three parts: bounding box regression loss, confidence loss, and classification loss.

Equalization cross entropy loss function and focal loss function are modified on the basis of the standard cross entropy loss function. The standard cross entropy loss function is shown as follows:

$$CE(p, y) = \begin{cases} -\log(p) & \text{if } (y = 1) \\ -\log(1 - p) & \text{if } (y = 0) \end{cases}, \tag{10}$$

where p is the predicted probability of the sample in this category and y is the sample label.

Equilibrium cross entropy loss function is based on the standard cross entropy loss function by multiplying a coefficient β to balance the weight of positive and negative samples as shown in the following formula:

$$CE(p, y) = \begin{cases} -\beta \log(p) & \text{if } (y = 1) \\ -(1 - \beta)\log(1 - p) & \text{if } (y = 0) \end{cases} \tag{11}$$

Focal loss is similar to balanced cross entropy. In order to improve the imbalance between positive and negative samples, weight α is introduced to improve the accuracy. Meanwhile, weight $(1-p)^{\gamma}$ is introduced to adjust the weight of difficult and easy samples, as shown in the following formula:

$$FL = \begin{cases} -\alpha(1 - p)^{\gamma}\log(p) & \text{if } (y = 1) \\ -(1 - \alpha)(p)^{\gamma}\log(1 - p) & \text{if } (y = 0) \end{cases}, \tag{12}$$

where α is the coefficient of the number of positive and negative samples and γ is a modulation factor that the greater the γ, the lower the contribution of simple sample loss. Focal loss acts on all prediction boxes during training, for both super parameters α and γ; in general, when α increase, γ should be appropriately reduced. In the experiment, the best effect was obtained when α of 2 and γ of 0.25 were selected.

In this paper, the focal loss function is used to replace the standard cross entropy function in the confidence loss in the YOLOv4 loss function; the category loss function in the YOLOv4 loss function is replaced by a balanced cross entropy loss function; the bounding box regression loss function remains unchanged. The improved loss function expression is shown as follows:

Figure 7: Unbalanced samples.

$$
\begin{cases}
L_{ciou} = \sum_{i=0}^{S^2}\sum_{j=0}^{B} I_{i,j}^{obj}\left[1 - IOU + \frac{\rho^2\left(b, b^{gt}\right)}{c^2} + \frac{16/\pi^4\left(\arctan w^{gt}/h^{gt} - \arctan w/h\right)^4}{1 - IOU + 4/\pi^2\left(\arctan w^{gt}/h^{gt} - \arctan w/h\right)^2}\right], \\
L_{\text{conf}} = \sum_{i=0}^{S^2}\sum_{j=0}^{B} I_{i,j}^{obj}\left[\widehat{C}_i \log\left(C_i\right) + \left(1 - \widehat{C}_i\right)\log\left(1 - C_i\right)\right], \\
-\lambda_{\text{noobj}} \sum_{i=0}^{S^2}\sum_{j=0}^{B} I_{i,j}^{\text{noobj}}\left[\widehat{C}_i \alpha\left(1 - C_i\right)^{\gamma}\log\left(C_i\right) + \left(1 - \widehat{C}_i\right)\left(1 - \alpha\right)\left(C_i\right)^{\gamma}\log\left(1 - C_i\right)\right], \\
L_{\text{class}} = -\sum_{i=0}^{S^2} I_{i,j}^{obj} \sum_{c\hat{I}\text{classes}}\left[\widehat{p}_i(c)\beta\ \log\left(p_i(c)\right) + \left(1 - \widehat{p}_i(c)\right)(1 - \beta)\log\left(1 - p_i(c)\right)\right], \\
\text{LOSS} = L_{ciou} + L_{\text{conf}} + L_{\text{class}},
\end{cases}
\tag{13}
$$

where S^2 represents the number of grids; B represents the number of prediction boxes in each cell; $I_{i,j}^{obj}$ and $I_{i,j}^{noobj}$ indicate whether there is an object in the i^{th} prediction box of the j^{th} unit; if there is an object, both take 1 and 0, respectively, and if there is no object, both take 0 and 1, respectively; λ^{noobj} represents the confidence penalty weight when the target is not included; ρ represents Euclidean distance; c represents the diagonal distance of the smallest closure area that contains both the prediction box and the ground truth box; b represents the canter coordinates of the prediction box; $w$and h, respectively, represent the width and height of the prediction box; b^{gt} is the canter coordinates of the ground truth box; w^{gt} and h^{gt}, respectively, represent the width and height of the ground truth box; C represents the true confidence of the target in the i^{th} grid; $\widehat{C}_i$ represents the reliability of the prediction box of the target in the i^{th} grid; pi(c) represents the category probability value of the ground truth box; and $\widehat{p}_i(c)$ represents the category probability value of the prediction box.

According to the experimental data in Table 2, it can be found that after focal loss improved YOLOv4's loss function, the average accuracy of the model was significantly improved.

Table 2: Comparison of experimental results of different optimization methods.

NMS	DANet	Improved NMS	Improved loss function	mAP (%)
√				90.8
√	√			93.7
		√		92.4
√			√	92.0
√	√		√	93.9
	√	√	√	94.7

Figure 8: Detection target: (a) insulator, (b) insulator defect, (c) spacer, (d) shock hammer, and (e) bird nest.

3. Model Training and Experimental Results Analysis

3.1. The Experimental Data. The drone aerial image used in this article was provided by a power grid in eastern Mongolia. A total of 2600 images were collected. The experiment detects five types of targets in the transmission line: insulators, insulator defects, bird nests, spacer, and shock hammers. The various targets are shown in Figure 8.

LabelImg is used to label the collected 2600 aerial images, and a VOC data set is made. 85% of the data (2210 images) were randomly selected as the training set and the remaining 15% (390 images) as the test set. The resolution of each original image is 4000×3000. Due to the limitation of computing resources, this article preprocesses the size of all training set and test set sample images so that the resolution of each image after processing is 608×608. The numbers and labels of the above five detection targets are shown in Table 3.

3.2. The Experimental Parameters. The experimental environment configuration of this article is as follows: processor model is i9-10900K@3.7 GHz; graphics card is NVIDIA 2∗RTX3090 24 GB GDDR6; operating system is

Table 3: Number of samples.

Label	Number
Insulator	5121
Insulator defect	506
Spacer	3530
Shock hammer	1570
Bird nest	158

Table 4: Experimental parameters.

Parameter name	Parameter value
CPU	2∗24 GB
Input image size	608∗608
DIoU	0.6
Momentum factor	0.9
Epoch	3000
First stage learning rate	0.001
Second stage learning rate	0.0001

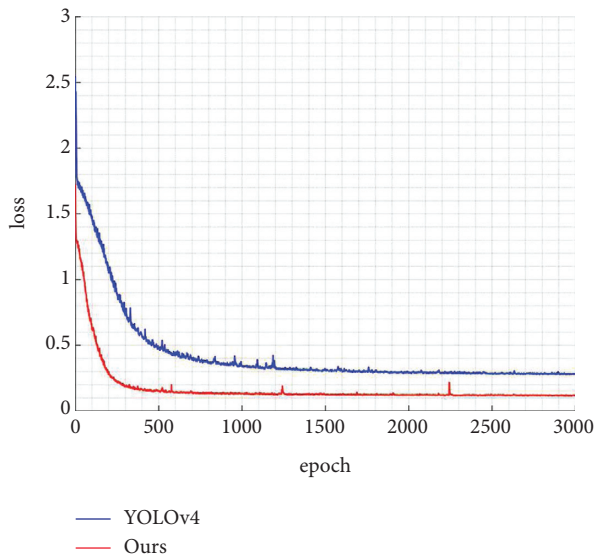

Figure 9: Loss curve.

Ubuntu18.04: deep learning framework Pytorch1.81. The experimental parameters are shown in Table 4.

3.3. Experimental Results and Analysis. The loss curve of the model is shown in Figure 9. The blue curve represents the change in the loss value of YOLOv4, and the red curve represents the change in the loss value of the improved model.

According to Figure 9, it can be found that the initial loss value of YOLOv4 is 2.55 and the initial value of the improved model is 1.73. With the continuous increase in epoch, the loss value is constantly decreasing. After 470 rounds of training, the loss value of YOLOv4 is reduced to about 0.5, and the loss value is finally stabilized at about 0.28. After 130 rounds of improved model training, the loss value dropped to about 0.5, and the loss value finally stabilized at about 0.12. Through the comparison of experimental results, it can be found that the loss value of the improved model decreases faster and the loss value is lower.

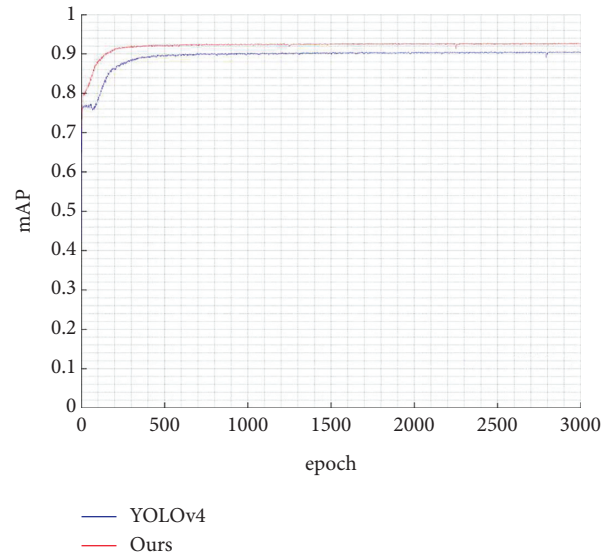

Figure 10: mAP value curve.

Table 5: Various target detection results.

Type	Precision	AP
Insulator	0.960	0.962
Insulator defect	0.950	0.948
Spacer	0.936	0.940
Shock hammer	0.912	0.908
Birdhouse	0.974	0.978
mAP	0.947	

The mean average precision curve of all categories of the model is shown in Figure 10. The blue curve represents the change in mAP (mean average precision) value of YOLOv4, and the red curve represents the change in mAP value of the improved model.

According to Figure 10, it can be seen that with continuous increase in model training rounds, the mAP value of YOLOv4 continues to rise, eventually reaches about 0.9. The mAP value of the improved model can finally reach about 0.92, with an increase of 2%. By comparing the loss value and the mAP value, it can be found that the improved YOLOv4 model has improved recognition accuracy and recognition effect compared with the original YOLOv4 model.

The detection results of the improved algorithm for various targets in the test set are shown in Table 5.

The experimental results of different optimization methods based on YOLOv4 are shown in Table 2.

The detection results of various targets by YOLOv4 are shown in Figure 11. The detection results of various targets by the improved algorithm are shown in Figure 12.

Figure 11: YOLOv4 algorithm detection results.

Comparing Figure 11 with Figure 12, it can be found that YOLOv4 has lower recognition accuracy for various types of targets in different backgrounds and misses detection of small targets and occluded targets. The improved algorithm can accurately identify various targets in different backgrounds.

3.4. Comparison of Different Algorithms. In order to verify the advantages of the proposed algorithm in multitarget detection of transmission lines, four traditional algorithms, such as Faster RCNN, SSD, RetinaNet, and YOLOv4, were selected for comparison. The loss curves of various models are shown in Figure 13.

The detection results of the above various algorithms are shown in Table 1.

According to the detection results in Table 1, it can be found that Faster RCNN has the highest mAP, which can reach 95.2%, but the detection speed is the lowest. YOLOv4 has the highest detection speed, but the accuracy is low. The detection accuracy of the algorithm proposed in this paper is slightly lower than that of Faster RCNN, but the detection speed is much higher than it. Taking the detection accuracy and real-time performance into account, it is more suitable for transmission line target detection.

Figure 12: Improved algorithm detection results.

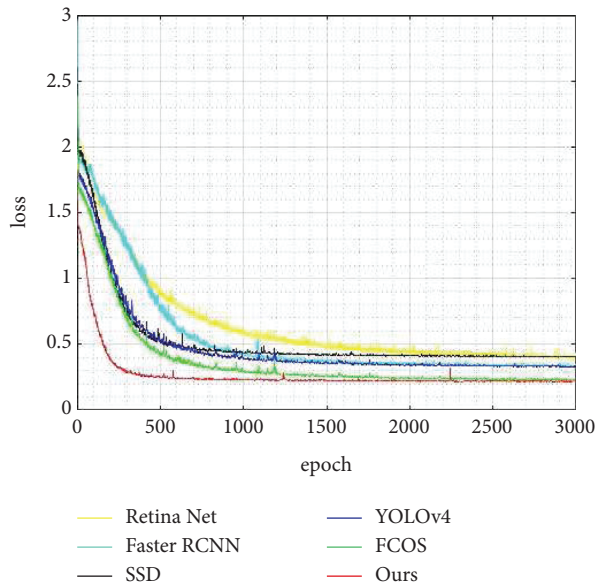

Figure 13: Loss value comparison.

4. Conclusions

(1) In order to solve the impact of the complex background of aerial images of transmission lines on target recognition, the fusion of DANet and YOLOv4 is adopted which can improve the detection accuracy by 3.7%

(2) In this paper, Gaussian function is used to optimize the nonmaximum suppression of the model, which can improve the detection accuracy of the occluded target

(3) This paper uses the focal loss function and the balanced cross entropy function to improve the loss function of YOLOv4, which reduces the imbalance between the background and the detection target as well as the impact of the imbalance of the sample on the target detection and improves the accuracy of the detection

Data Availability

The test data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work. The authors declare that they do not have any commercial or associative interest that represents conflicts of interest in connection with the work submitted.

Acknowledgments

This work was supported by Liaoning Provincial Department of Education Scientific Research Funding Project (LJ2019JL013 and LJ2020JCL020) and Liaoning University of Engineering and Technology Discipline Innovation Team Funding Project (LNTU20TD-29).

References

[1] L. Yun, "Common problems in the operation of the high voltage transmission line and its maintenance," *Journal of communication power supply technology*, vol. 37, no. 1, pp. 273-274, 2020.

[2] Y. Sui, P. Ning, and P. Niu, "A review of power inspection technology of mounted unmanned aerial vehicle for overhead transmission lines," *Power Grid Technology*, vol. 45, no. 9, pp. 3636–3648, 2021.

[3] R. Girshick, J. Donahue, and T. Darrell, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587, Columbus Ohio, USA, June 2014.

[4] R. Girshick, "Fast r-cnn," in *Proceedings of the 2015 IEEE International Conference on Computer Vision(ICCV)*, pp. 1440–1448, IEEE, Santiago , Chile, December 2015.

[5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[6] J. Dai, Y. Li, K. He, and R. Fcn, "Object detection via region-based fully convolutional networks," in *Proceedings of the Advances in Neural Information Processing Systems*, December 2016.

[7] W. Liu, D. Anguelov, D. Erhan et al., "SSD: ss detector," in *Proceedings of the Computer Vision - ECCV 2016*, pp. 21–37, Amsterdam, The Netherlands, October 2016.

[8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection[C]," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern*, pp. 779–788, IEEE, Las Vegas, NV, USA, June 2016.

[9] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6517–6525, IEEE, Honolulu,HI, USA, July 2017.

[10] J. Redmon and A. Farhadi, "YOLOv3: an incremental improvement," 2018, https://arxiv.org/abs/1804.02767.

[11] Z. Zhao, H. Qi, Y. Qi, K. Zhang, Y. Zhai, and W. Zhao, "Detection method based on automatic visual shape clustering for pin-missing defect in transmission lines," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 9, pp. 6080–6091, 2020.

[12] X. Tao, D. Zhang, Z. Wang, X. Liu, H. Zhang, and D. Xu, "Detection of power line insulator defects using aerial images analyzed with convolutional neural networks," *IEEE Transactions on Systems, Man, and Cybernetics:. Systems*, vol. 50, no. 4, pp. 1486–1498, 2018.

[13] Y. Liu, X. Ji, S. Pei et al., "Research on automatic location and recognition of insulators in substation based on YOLOv3," *High Voltage*, vol. 5, no. 1, pp. 62–68, 2020.

[14] Q. Wang and B. Yi, "Recognition of insulator defects in aerial images based on Gaussian YOLOv3[J]," *Progress in Laser and Optoelectronics*, vol. 58, no. 12, pp. 254–260, 2021.

[15] J. Fu, J. Liu, H. Tian et al., "Dual attention network for scene segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3146–3154, Long Beach, CA, USA, 2019.

[16] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: optimal speed and accuracy of object detection," 2020, https://arxiv.org/abs/2004.10934.

[17] T. Y. Lin, P. Goyal, and R. Girshick, "Focal loss for dense object detection," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2999–3007, Venice, Italy, October 2017.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[19] T. Y. Lin, P. Dollar, and R. Girshick, "Feature pyramid networks for object detection," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE Computer Society, Hawaii, United states, July 2017.

[20] A. Neubeck and G. Van, "Efficient non-maximum suppression," in *Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)*, Auguest 2006.

[21] Y. He, C. Zhu, J. Wang, M. Savvides, and X. Zhang, "Bounding box regression with uncertainty for accurate object detection," in *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[22] D. Xu, L. Wang, and F. Li, "Research Summary of Typical Target Detection Algorithms for Deep learning," *Computer Engineering and Applications*, vol. 57, no. 8, pp. 10–215, 2021, https://iopscience.iop.org/article/10.1088/1742-6596/1757/1/012003/pdf.