Hindawi
Mobile Information Systems
Volume 2022, Article ID 7408265, 12 pages
https://doi.org/10.1155/2022/7408265

Research Article

Corrosion Detection Method of Transmission Line Components in Mining Area Based on Multiscale Enhanced Fusion

Jun Wu, Yingxin Sun, and Xinliang Wang

Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China

Correspondence should be addressed to Xinliang Wang; wangxl@hpu.edu.cn

Received 5 August 2022; Revised 1 September 2022; Accepted 18 September 2022; Published 4 October 2022

Academic Editor: Yuexia Zhang

Component corrosion is one of the potential safety hazards in transmission lines in mining areas. In order to solve the problem of poor detection accuracy caused by the large proportion of small targets and complex background in the current distant view corrosion inspection task by UAV, we propose a PWR-YOLOV5 detection method for corrosion components based on the YOLOV5 algorithm. Firstly, a new feature fusion network, WA-PANet, is reconstructed on the basis of the path aggregation network (PANet) to make full use of the features at different stages and advance the detection accuracy of small targets in distant view by deepening the process of feature fusion and introducing the skip layer connections and adaptive feature fusion factors. Secondly, the pyramid split attention (PSA) module is introduced into the deep layers of the network to highlight the feature expression of corrosion targets and enhance the ability to detect pixel-level objects. Then, we construct a receptive feature enhancement network (RFENet), which can heighten the feature fusion effect of the WA-PANet and alleviate the problem of the feature expression ability weakening due to the fusion of different receptive field features. Finally, the EIoU Loss is adopted to optimize the loss function and improve the positioning accuracy of the bounding box. The experimental results show that the mAP of the PWR-YOLOV5 algorithm can reach up to 95.37%, which is 5.22% higher than YOLOV5, and the detection speed is 64.9FPS. Compared with the algorithms such as YOLOV4, Faster *R*-CNN, and YOLOX, the improved algorithm has better overall detection performance for the corrosion components of transmission lines in the mining area.

1. Introduction

At present, mineral security has risen to the level of a national strategy [1]. As an important branch of the national power system, the maintenance of the power supply system in mining areas plays a key role in the safety of mineral production work. Transmission lines in mining areas are mostly erected between mountains, rivers, and hills, and the climate is humid. Antivibration hammers, insulators, and other metal components are prone to rust on account of exposure to a harsh environment for long term and may even cause power supply faults such as components falling and line breakage, which will affect the normal operation of the power system in the mining area and seriously threaten the mineral security [2]. Therefore, it is necessary to effectively recognize and detect the corrosion flaws of transmission line components and timely find rust spots and repair corrosion problems, so as to ensure the safe and stable running of the power supply equipment in mining areas.

With Jones in [3] and Araar et al. in [4] introducing small unmanned aerial vehicles (UAV) into the inspection task of transmission lines, the inspection method of "UAV inspection + manual processing" appeared. A UAV was used instead of manual work to collect images, which could greatly reduce front-end labor costs. However, the way of manual detection was greatly influenced by subjective consciousness, resulting in serious missed and false detection [5]. The rapid development of computer technology had realized the combination of computational pattern recognition and power supply system inspection tasks, and then, a new inspection method of "UAV inspection + image processing" came into being [6]. In the task of detecting the corrosion components by using image processing technology, Recky and Leberl [7] used color space combined with a

k-nearest neighbor window to realize the detection of corrosion defects. A detection method for corrosion flaws of antivibration hammers was given in [8] by introducing histogram equalization, median filtering, morphological processing, and the RGB color model. The authors of [9] studied the color and texture features of corrosion images and applied the HSI color model and grey level co-occurrence matrix to identify corroded areas of images. Huang et al. [10], respectively, adopted edge features' enhancement algorithms such as local difference and anisotropic Gaussian kernel directional derivative, and threshold segmentation and morphological processing techniques on grayscale images to obtain the corrosion area ratio and color shadow index of antivibration hammers, which realized the classification of corrosion grade. On the basis of texture features of images, the authors in [11] also combined Fourier transform and multiple color models to distinguish corrosion areas. The methods mentioned above improve the accuracy of corrosion detection to a certain extent, but they are only applicable to detecting corrosion components with a relatively simple background, sparse target distribution, and obvious edge features.

In recent years, machine learning has developed rapidly owing to the successive optimization of calculation capacity and computational overhead. In [12], the authors extracted color, gradient amplitude, and direction histogram, respectively, by letting the aggregation channel feature algorithm to build a multiscale ACF pyramid and also integrated it with the AdaBoost classifier, Graph Cuts algorithm, and RGB color model to judge whether the antivibration hammers were corroded. The algorithm achieved higher detection accuracy, but it requested the way of manual design to extract features, which was complicated in steps and heavy in workload. Keeping up with the boom of object detection technology in deep learning, Petricca et al. in [13] led a convolutional neural network into the field of corrosion detection, providing a new thought for corrosion component recognition. The authors of [14] first used Retinex, an image enhancement algorithm, to decrease the interference of light and shadow on the corrosion color and then employed FPN and RPN structures to redesign the Faster-RCNN to achieve the defects classification and position regression of anti-vibration hammers, which optimized positioning accuracy of the algorithm to a certain degree. Combined with the attention mechanism, a lightweight corrosion targets' detection method for power equipment based on the SSD was presented in [15], realizing the identification of rusted areas with fewer parameters, but the average precision was only 71.35%. In [16], the authors advanced an attention-guided multitask convolutional neural network and connected it with the RPN structure to identify the corrosion degree and abnormal state of power line components. The methods mentioned above lack specific classification, and the shapes and appearances of different components are disparate, resulting in a high detection error rate.

To sum up, there are few studies on the detection of specific corrosion components of transmission lines at present, and the detection effect of related corrosion studies is not satisfactory. So, we put the YOLOV5 [17] to the task of detecting two corrosion components of transmission lines in mining areas, including antivibration hammers and insulators. According to the problems that there are a high proportion of small targets and a great many background interferents all caused by long-range shootings in image acquisition by UAV, we propose four improvements and optimizations as follows:

(i) The weight adaptive path aggregation network (WA-PANet) is constructed. On the basis of the PANet, we deepen the process of feature fusion and inlet skip layer connections and adaptive feature fusion factors to enhance the detection accuracy of different scale objects.

(ii) We introduce the PSA mechanism to fuse context information of different scales and meanwhile generate pixel-level attention for the targets, so as to highlight the feature expression of small corrosion targets.

(iii) The RFENet is built to strengthen the fusion effect of the WA-PANet and alleviate the weakening of the feature expression ability induced by feature fusion at different stages.

(iv) The loss function of bounding box regression adopts the EIoU Loss, which can improve the positioning accuracy and convergence rate of the network.

2. Materials and Methods

2.1. Data Acquisition and Processing. The data of transmission line components in the mining area required for the experiments in this study are provided by the Hemei Group Power Supply Department. We use the Pr software to extract frames from the video shot by the UAV and filter out a large number of similar and background pictures. There are two kinds of image resolutions 5184×3888 pixels and 5472×3078 pixels, respectively. On account the fact that the data given above covers all components of overhead transmission lines, we further screen out a total of 2705 images containing antivibration hammers and insulators and then adopt the LabelImg, a deep learning target annotation tool, to mark the objects. The labels are set to FangRust (corrosion antivibration hammers), Fang_NoRust (noncorrosion antivibration hammers), Jue_Rust (corrosion insulators) and Jue_NoRust (noncorrosion insulators), and the tagging format is VOC format. The data processing part applies the adaptive image scaling method to uniformly scale the read images of different scales to the network input size and accomplishes online data augmentation through random clipping, mosaic data enhancement, etc., which enhances the robustness of the model and improves the detection performance of the algorithm for different scale targets, especially the small ones.

2.2. PWR-YOLOV5 Network. Although the YOLOV5 algorithm shows a good performance in generic object detection tasks, the color of corrosion is easily confused with background items such as dead leaves and dust and is

also greatly affected by the intensity of light. In addition, the shooting distance of the UAV is not flexible to control, leading to many small targets in the distant view, all of which have a poor influence on the recognition of rusted components. In order to adapt to the corrosion targets' detection task in the actual scene, in this study, we first redesign the feature fusion network WA-PANet by deepening the feature fusion process and setting skip layer connections and learnable feature fusion arguments, which not only retains more details but also improves the feature fusion effect. Then, the PSA mechanism is pulled into the deep layers of the network, in which the softmax is used to adaptively fuse the spatial features of different scales and channel attention weights to generate pixel-level attention to the objects. Next, we apply bottleneck structures and dilated convolutions to construct the feature enhancement network RFENet to capture the multiscale features under different receptive fields in order to strengthen the fusion effect of the WA-PANet and enhance the detection accuracy of corroded targets at different scales. Finally, the bounding box regression loss function is optimized by the EIoU Loss, which solves the problem of ambiguous definition of aspect ratio loss in CIOU Loss and advances the positioning accuracy of the network. Based on the innovations mentioned above, the PWR-YOLOV5 detection algorithm for corrosion components is proposed. The structure of the PWR-YOLOV5 is shown in Figure 1.

2.2.1. Weight Adaptive Path Aggregation Network. The YOLOV5 algorithm uses the PANet for feature fusion, and its structure is shown in Figure 2(a). In this module, the method of bidirectional fusion is deployed to integrate deep semantic information and shallow location information, which promotes detection accuracy to a certain extent. However, this fusion method does not distinguish the feature information at different stages, inducing negative feature fusion results. For the purpose of solving the problem and boosting the detection performance of the rusted small targets, on the basis of the PANet, we make the process of feature fusion deeper, and meanwhile, import skip layer connections and adaptive feature fusion factors [18] to establish the weight adaptive path aggregation network. The specific structure of the WA-PANet is shown in Figure 2(b), where P_i represents the ith feature produced in the backbone and F_i and N_i are the intermediate features generated during the fusion procedure.

The WA-PANet feature fusion network consists of two branches. The one top-down branch is used to transmit the powerful semantic information and perform the fusion of deep and shallow features, where the fusion mode of channel cascade is applied to preserve more feature information. In this study, we add an up-sampling operation to form the feature map F_2, which is fused with the feature map P_2 yielded in the backbone to generate the large-scale feature N_2 with rich semantic and location information so as to improve the performance of the network to identify small corroded targets. The other bottom-up branch is employed to transfer detailed information with the purpose of enhancing the ability to locate objects at different scales. For the sake of obtaining refined features, the skip layer connections from input to output are set at nodes N_3, N_4, and N_5, respectively. Furthermore, due to the multiple input features of these three nodes coming from different stages of the network and contributing differently to the fusion results, we also introduce a learnable feature factor for each branch of each node so that the algorithm adaptively learns the importance of different stage features in the training process. The features of each node of the WA-PANet can be expressed as the following equations (1)–(5):

$$F_5 = Conv(P_5), \tag{1}$$

$$F_i = Conv(\mathrm{Concat}(P_i, \mathrm{Upsample}(F_{i+1}))), i = 2, 3, 4, \tag{2}$$

$$N_2 = Conv(F_2), \tag{3}$$

$$N_i = Conv\left(\frac{w_1 \times P_i + w_2 \times F_i + w_3 \times Resize(N_{i-1})}{w_1 + w_2 + w_3 + \varepsilon}\right), i = 3, 4, \tag{4}$$

$$N_5 = Conv\left(\frac{w_1 \times F_5 + w_2 \times Resize(N_4)}{w_1 + w_2 + \varepsilon}\right), \tag{5}$$

where *Conv* refers to a series of convolution operations involved in feature processing, *Upsample* stands for the nearest neighbor interpolation up-sampling, *Concat* means that two features from different stages carry out splicing and fusion on the channel dimension, that is, channel cascade, and *Resize* means to adjust the size of the feature maps in the two dimensions of space and channel. And w_i is a learnable parameter, which is multiplied by the input feature of the corresponding branch at each node, so the larger the value of w_i is, the greater the influence of the branch on the fusion results is, and ε is a constant, much less than 1, which is set to prevent the denominator from being 0.

2.2.2. Pyramid Split Attention Mechanism. There are a large number of interruptions in the actual scene, and the background is complicated. At the same time, with the deepening of the network, the features of small objects with fewer pixels are gradually blurred, and the location information is also less and less obvious. In an effort to effectively suppress the complex background information and overcome the influence of light intensity, the pyramid split attention mechanism [19] is introduced into the deep layers of the network to promote the detection accuracy for corroded components. The implementation of the PSA module is shown in Figure 3.

First of all, we utilize the split and concatenation (SPC) structure to obtain feature information of different scales along the channel direction, and the SPC composition is demonstrated in Figure 4. The module divides the input feature maps X into S parts in the channel dimension, denoted as $[X_1, X_2, \ldots, X_{S-1}]$, so each part has the same number of channels $C' = C/S$, and the feature of the ith

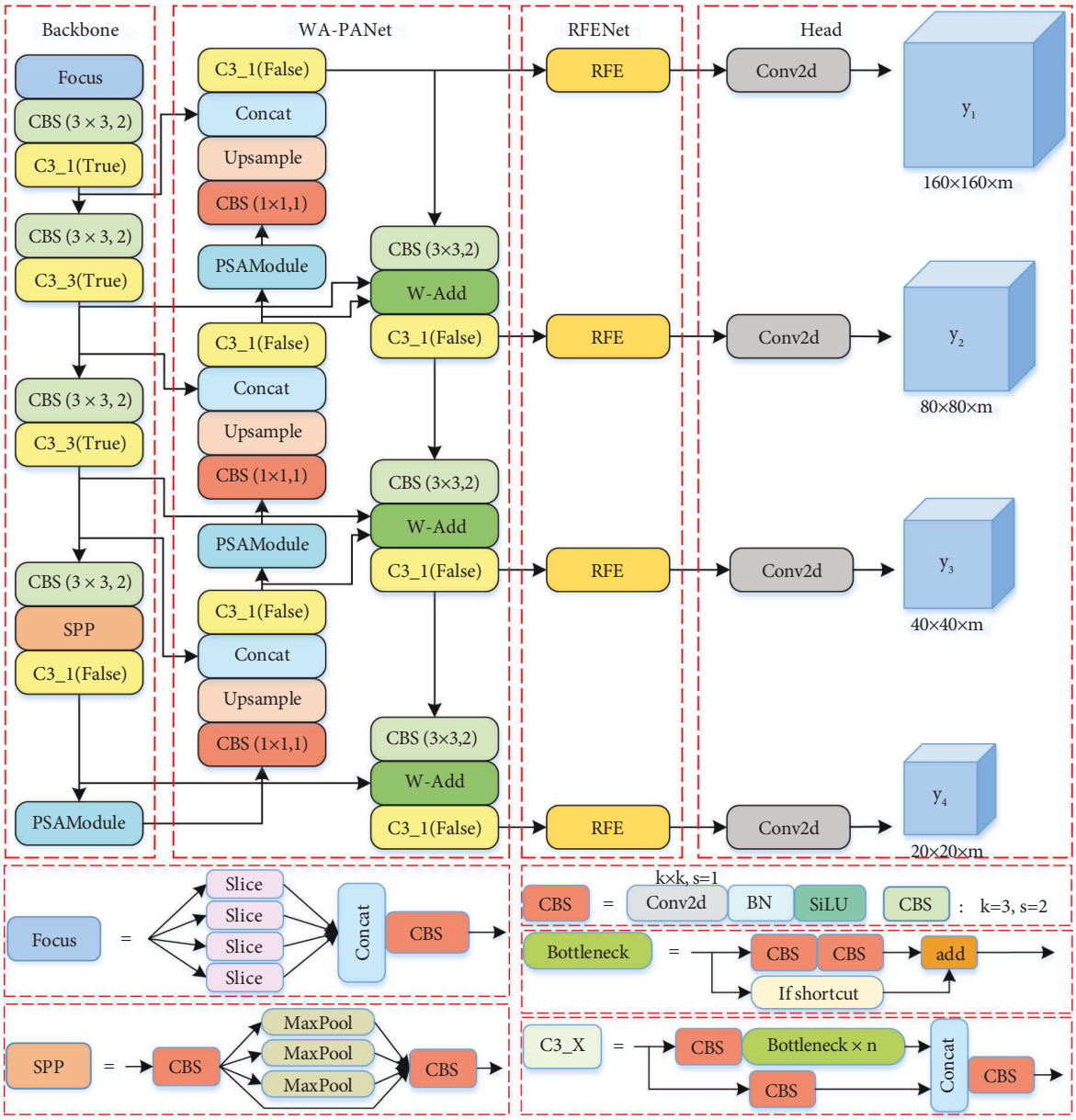

Figure 1: The PWR-YOLOV5 network structure. P is the PSA module; W presents WA-PANet; and R is RFENet.

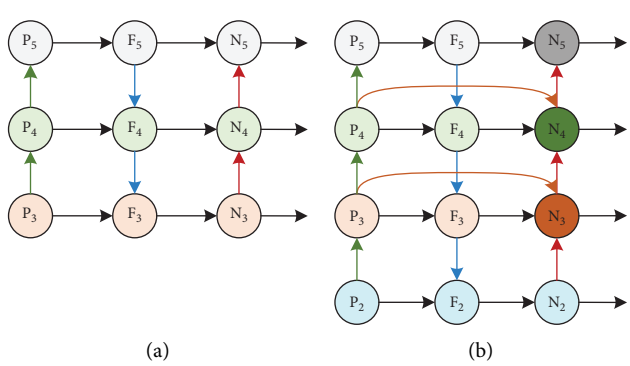

Figure 2: (a) The path aggregation network structure. (b) The weight adaptive path aggregation network structure.

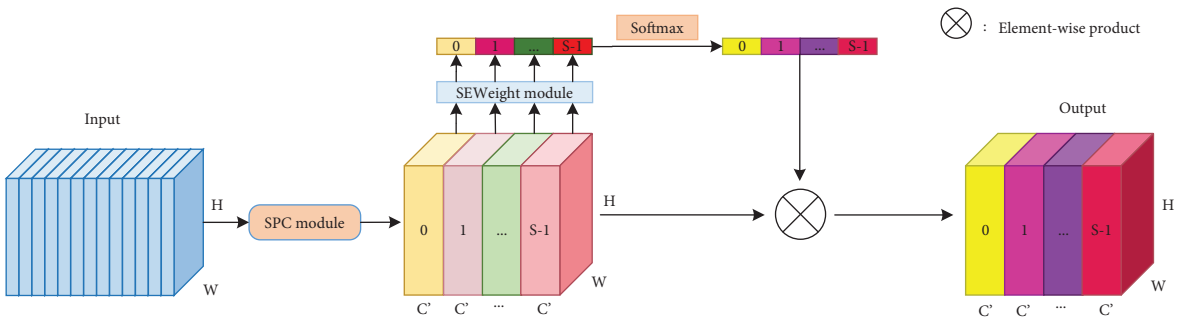

Figure 3: The pyramid split attention structure.

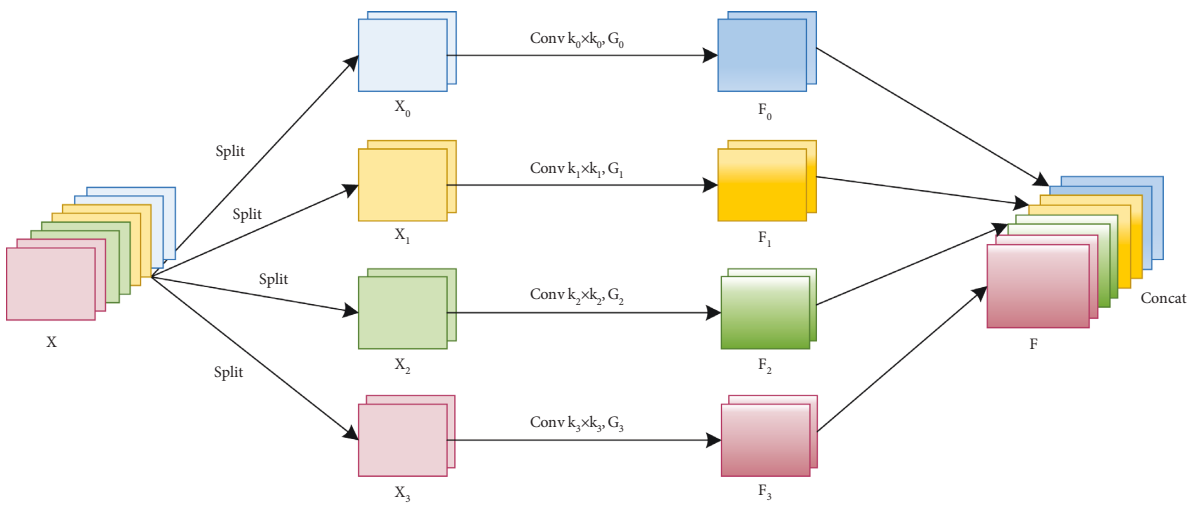

Figure 4: The split and concat structure.

branch can be represented as $X_i \in R^{C' \times H \times W}, i = 0, 1, \ldots, S-1$. In this study, the value of S is 4. After being divided in this way, the input tensors are processed in parallel by using convolution kernels of different scales, so as to extract spatial information from feature maps of each branch and gain features of different receptive fields and depths. However, with the augmentation of the convolution kernel size, the number of parameters also increases significantly. In order to save computational overhead, the group convolution is inlet; moreover, a new rule is also designed for the selection of the grouping number. The relationship between the size of multiscale convolution kernels and the number of groups can be represented as the following formula:

$$G = 2^{(k-1)/2}, \tag{6}$$

where K is the size of the convolution kernel and G is the number of groups. In particular, the number of groups is defaulted to 1 when $K = 3$. The generating process of the multiscale features is given in (7), and then, the different scale features produced by each branch are cascaded in channel to obtain the feature map F in the following formula:

$$F_i = Conv(k_i \times k_i, G_i)(X_i), i = 0, 1, 2, \ldots, S-1, \tag{7}$$

$$F = \text{Concat}([F_0, F_1, \ldots, F_{S-1}]), \tag{8}$$

where $k_i = 2 \times (i+1) + 1$ denotes the size of convolution kernel applied in the ith branch, and the grouping number of convolution operation in the ith branch is $G_i = 2^{(k_i-1)/2}$, and $F_i \in R^{C' \times H \times W}$ signifies the feature of different receptive fields emerged on each branch, and $F \in R^{C \times H \times W}$ is the complete multiscale feature maps acquired through the SPC module.

Secondly, the SEWeight [20] is used to pick up the channel attention information of each branch feature, and the attention weight vector of each branch channel is procured by (9), in which $Z_i \in R^{C' \times 1 \times 1}$ shows the channel attention weights obtained from the different scale features F_i:

$$Z_i = \text{SEWeight}(F_i), i = 0, 1, 2, \ldots, S-1. \tag{9}$$

Then, the soft attention method is employed to handle the channel attention weight vector Z_i to adaptively select the importance of multiscale spatial features across channels. The weight allocation pattern of soft attention is shown in formula (10), where Softmax is adopted to weighting fusion of the spatial and channel information of each branch to obtain the weight att_i, which contains the information of all positions in the space and the attention weight in the channel:

$$att_i = Softmax(Z_i) = \frac{\exp(Z_i)}{\sum_{i=0}^{S-1} \exp(Z_i)}, i = 0, 1, 2, \ldots, S-1. \tag{10}$$

Finally, the fused attention weight att_i of each branch is multiplied by the corresponding scale feature F_i to fetch the feature map Y_i with multiscale pixel-level channel attention in formula (11), and in the end, the multiscale refined output is acquired through the concatenation operation that can maintain the integrity of features, and the course can be denoted as follows:

$$Y_i = F_i \odot att_i, i = 0, 1, 2, \ldots, S-1, \tag{11}$$

$$\text{Out} = \text{Concat}\left(\left[Y_0, Y_1, \ldots, Y_{S-1}\right]\right), \tag{12}$$

where $\odot$ indicates pixel-wise multiplication. The PSA module can integrate multiscale spatial information and cross-channel attention through each split feature group, which can not only achieve the fusion of different scales of context information but also generate pixel-level attention to the targets. In this study, we put the PSA module into the last layer of the backbone and the top-down process of the WA-PANet to strengthen the suppression of complex interference information at a higher semantic level and highlight the feature expression effect of small corrosion targets.

2.2.3. Receptive Feature Enhancement Network. In the bottom-up course of the WA-PANet, features from different stages need to be fused. Owing to the receptive fields of different branch features are different, so the semantic information is dissimilar. The fusion of multiple feature maps at different semantic levels will greatly weaken the expression ability of multiscale features, which is not conducive to the detection results of the algorithm. Based on the problem, we use the bottleneck layers and dilated convolutions of different scales to construct the receptive feature enhancement network [21] that carries out enhanced extraction on the multibranch fusion outputs of the WA-PANet to advance the feature expression of rusted targets at various scales. The structure of the RFENet module is shown in Figure 5.

The module is a multibranch structure. Firstly, a bottleneck layer is adopted in each branch, namely, a 1×1 convolution for dimensionality reduction and a $n \times n$ convolution to achieve the extraction of different scale features, and then, a 3×3 dilated convolution with a dilation rate of n is followed to capture the feature information in a larger receptive field area, and ultimately, the concat and shortcut operations are applied to fuse the features of different receptive fields. With the purpose of compressing the number of parameters, we deploy two 3×3 convolutions, as well as 1×7 and 7×1 convolutions to replace 5×5 and 7×7 convolutions, respectively. In this study, the module is placed in the four output branches of the WA-PANet to establish the RFENet, which can intensify the effect of feature fusion and promote the recognition accuracy for corrosion objects of different scales.

2.2.4. EIoU Loss. The YOLOV5 algorithm applies the CIOU Loss as the loss function of bounding box regression, which takes into account three geometrical factors, including overlapping area, center points' distance, and aspect ratio. Given a prediction box B and a ground truth B^{gt}, the CIOU Loss can be defined as

$$\begin{cases} L_{\text{CIOU}} = 1 - IOU + \dfrac{\rho^2\left(b, b^{gt}\right)}{c^2} + \alpha\nu, \\ \nu = \dfrac{4}{\pi^2}\left(\arctan\dfrac{w^{gt}}{h^{gt}} - \arctan\dfrac{w}{h}\right)^2, \\ \alpha = \dfrac{\nu}{(1 - IOU) + \nu}, \end{cases} \tag{13}$$

where IOU is the ratio of the intersection area and the union area of the prediction box and the ground truth, b and b^{gt} represent the center of B and B^{gt}, respectively, $\rho(\cdot) = \|b - b^{gt}\|_2$ is the distance between b and b^{gt}, c is the diagonal length of the minimum bounding rectangle between the prediction box and the ground truth, and α and ν are used to reflect the similarity of the aspect ratio between the prediction box and the ground truth. In (13), ν reveals the difference between aspect ratios, rather than the real relationship between w and w^{gt} or h and h^{gt}. When w and h meet formula $\{(w = kw^{gt}, h = kh^{gt}) | k \in R^+\}$, the value of ν is 0, indicating that the length and width of the prediction box and the ground truth are completely matched. As a result, the loss of aspect ratio item is 0, and the bounding box regression process is blocked, which is inconsistent with reality. In addition, in the training process, the calculating process of ν for w and h back propagation to obtain the gradient is denoted in formula (14). According to the formula, we can get $\partial\nu/\partial w = -h/w\partial\nu/\partial h$, in which the signs of $\partial\nu/\partial w$ and $\partial\nu/\partial h$ are opposite. Thus, if one of these two variables (w or h) increases, the other will decrease, which prevents the reduction of the real difference between the prediction box and the ground truth:

$$\begin{cases} \dfrac{\partial\nu}{\partial w} = \dfrac{8}{\pi^2}\left(\arctan\dfrac{w^{gt}}{h^{gt}} - \arctan\dfrac{w}{h}\right) * \dfrac{h}{w^2 + h^2}, \\ \dfrac{\partial\nu}{\partial h} = -\dfrac{8}{\pi^2}\left(\arctan\dfrac{w^{gt}}{h^{gt}} - \arctan\dfrac{w}{h}\right) * \dfrac{w}{w^2 + h^2}. \end{cases} \tag{14}$$

Owing to the unclear definition of ν in the last item of L_{CIOU}, the convergence speed and positioning accuracy of the algorithm are limited. Therefore, we introduce the EIOU Loss [22], an optimization version of the CIOU Loss, to compute regression loss. It is defined as

$$\begin{aligned} L_{\text{EIOU}} &= L_{\text{IOU}} + L_{dis} + L_{asp} \\ &= 1 - IOU + \frac{\rho^2\left(b, b^{gt}\right)}{c^2} + \frac{\rho^2\left(w, w^{gt}\right)}{c_w^2} + \frac{\rho^2\left(h, h^{gt}\right)}{c_h^2}, \end{aligned} \tag{15}$$

where c_w and c_h are the length and width of the minimum bounding rectangle covering the prediction box and the ground truth, respectively. The EIOU loss also consists of

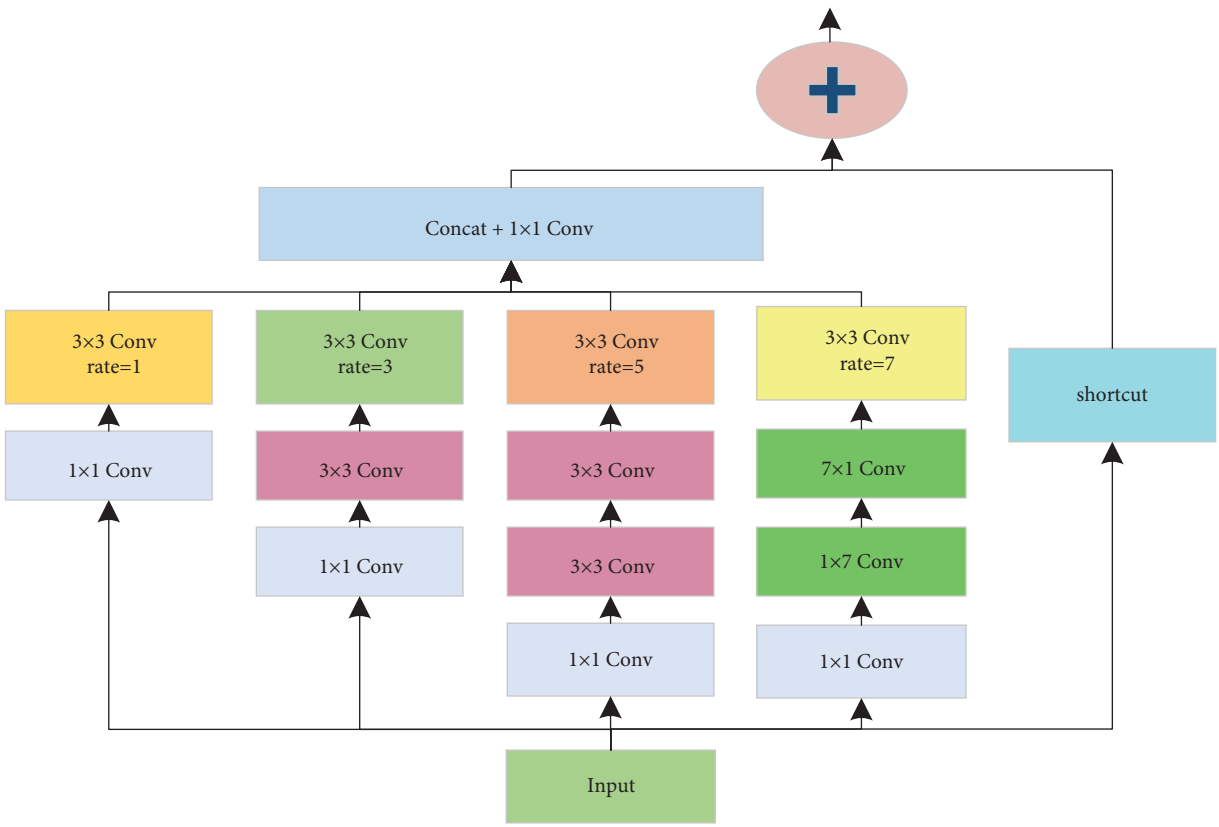

Figure 5: The feature enhancement network module structure.

three parts, the overlapping area loss L_{IOU}, the center points distance loss L_{dis}, and the length and width difference loss L_{aps}. Among them, L_{aps} can directly minimise the length-width gap between the prediction box and the ground truth so that the convergence speed and the location performance of the network are excellent.

3. Results and Discussion

3.1. Experimental Environment and Parameters' Setting. Based on the Ubuntu 18.04 operating system, this study employs the PyTorch deep learning framework and *Python* compiler language to carry out all the experiments. The specific experimental environment and training parameter settings of the improved algorithm are shown in Table 1.

3.2. Algorithm Evaluation Index. In order to verify the effectiveness of the innovative improved method in this study, the Average Precision (AP), the mean Average Precision (mAP), the Frames Per Second (FPS), and the Model Size (MS) are selected to evaluate model performance according to the positioning accuracy of the prediction boxes and the missed and false detection of the targets. The AP measures the precision and recall of a certain class, and its value is the area of the *P-R* curve. The larger the value is, the better the detection performance of the network for this kind of target. The mAP is the mean value of the AP of various categories, which is used to evaluate the overall detection accuracy of the model. The FPS reflects the detection speed; the greater the value is, the better the real time of the algorithm is. The MS refers to the amount of memory occupied by the algorithm, which represents the requirement for storage space.

3.3. Experimental Results and Comparative Analysis

3.3.1. Experimental Result of the PWR-YOLOV5. The self-made dataset of corrosion components is divided into a training set, a validation set, and a test set in a ratio of 8 : 1 : 1. In order to avoid the influence of video frame extraction on the distribution of the dataset, all images are shuffled before partitioning to advance the generalization of the model. In this study, the improved algorithm is trained and tested based on the above partitioning method. Figure 6 shows the *P-R* curve of the PWR-YOLOV5 network during testing, which is drawn from the precision and recall values under all confidence levels of various targets. The area below the curve indicates the average precision of each class, so the AP of FangRust, Fang_NoRust, Jue_Rust, and Jue_NoRust, respectively, reaches up to 96.88%, 95.02%, 95.61%, and 93.96%.

3.3.2. Comparison with YOLOV5. In this study, the YOLOV5 and the modified PWR-YOLOV5 are compared in the following four aspects: average precision, mean average accuracy, frames per second, and model size. The comparative results are recorded in Table 2.

Compared with YOLOV5, the AP of the proposed algorithm on FangRust, Fang_NoRust, Jue_Rust, and

Table 1: Experimental environment and parameter settings.

Experimental environment		Model parameter	
Parameters	Configuration	Parameters	Value
CPU	Intel(*R*) xeon(*R*) *E*5-2680 v4	Optimizer	SGD
GPU	NVIDIA GeForce GTX 3080	Initial learning rate	0.001
CUDA	11.0	Size of images	640 × 640
Python	3.8	Batch size	16
PyTorch	1.9.0	Epoch	350

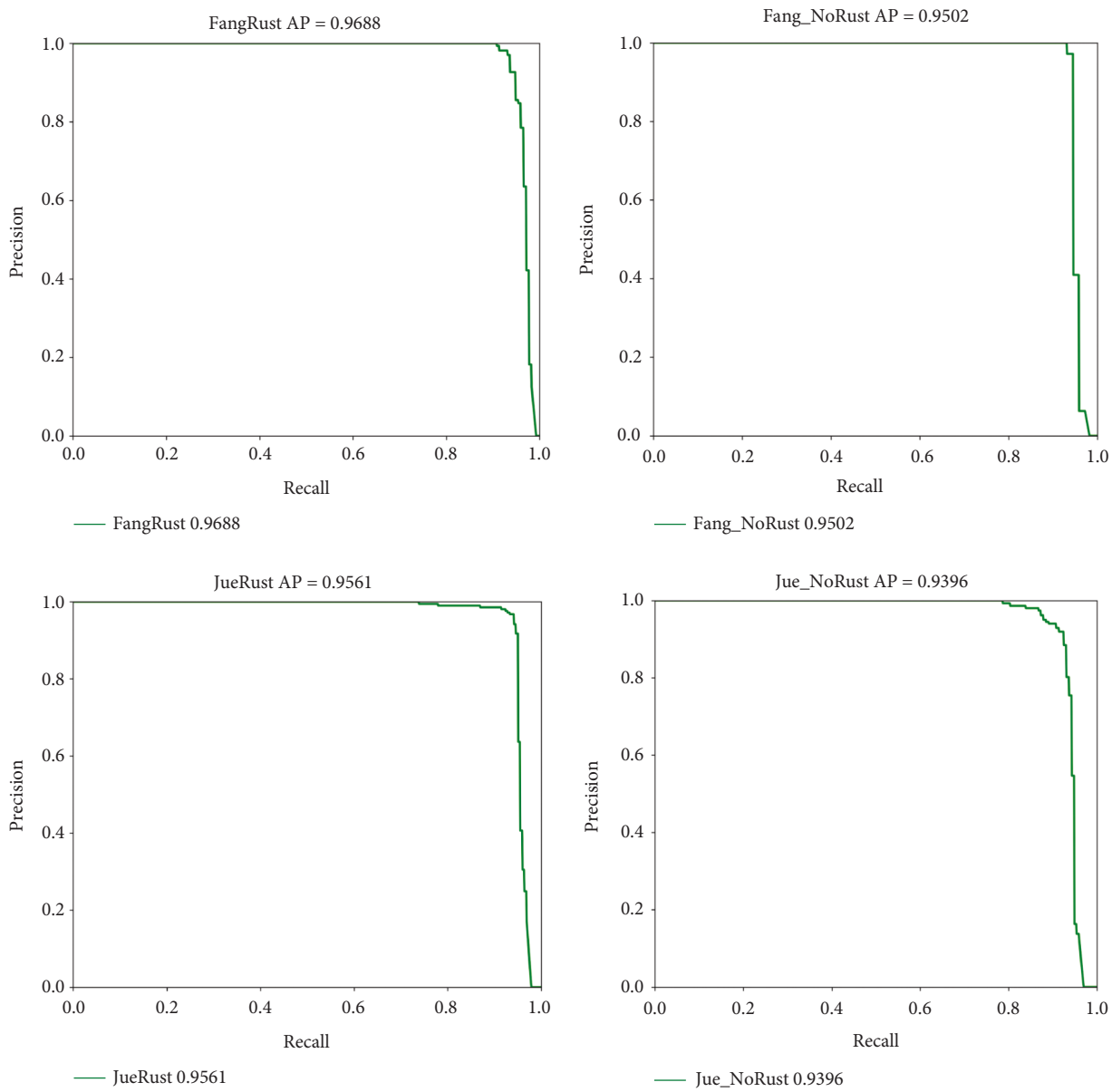

Figure 6: The *P*-*R* curve for four classes of the PWR-YOLOV5 algorithm.

Jue_NoRust is improved by 2.52%, 5.3%, 2.11%, and 10.93%, respectively, and the mAP is increased by 5.22%. The experimental data given above show that the improved method presented in this study can effectively advance the detection accuracy of various targets. However, the introduced PSA modules and RFENet in the PWR-YOLOV5 all increase the number of parameters, resulting in the model being inferior to the YOLOV5 in terms of detection speed and memory overhead.

The detection effect of the algorithm in real complex scenes is shown in Figure 7, in which the original images are on the left, and the YOLOV5 detection results are in the

Table 2: Performance comparison between the YOLOV5 and the improved PWR-YOLOV5.

Models	AP (%)				mAP (%)	FPS	MS (MB)
	FangRust	Fang_NoRust	JueRust	Jue_NoRust			
YOLOV5	94.36	89.72	93.50	83.03	90.15	**109.9**	**14.40**
Ours	**96.88**	**95.02**	**95.61**	**93.96**	**95.37**	64.9	26.65

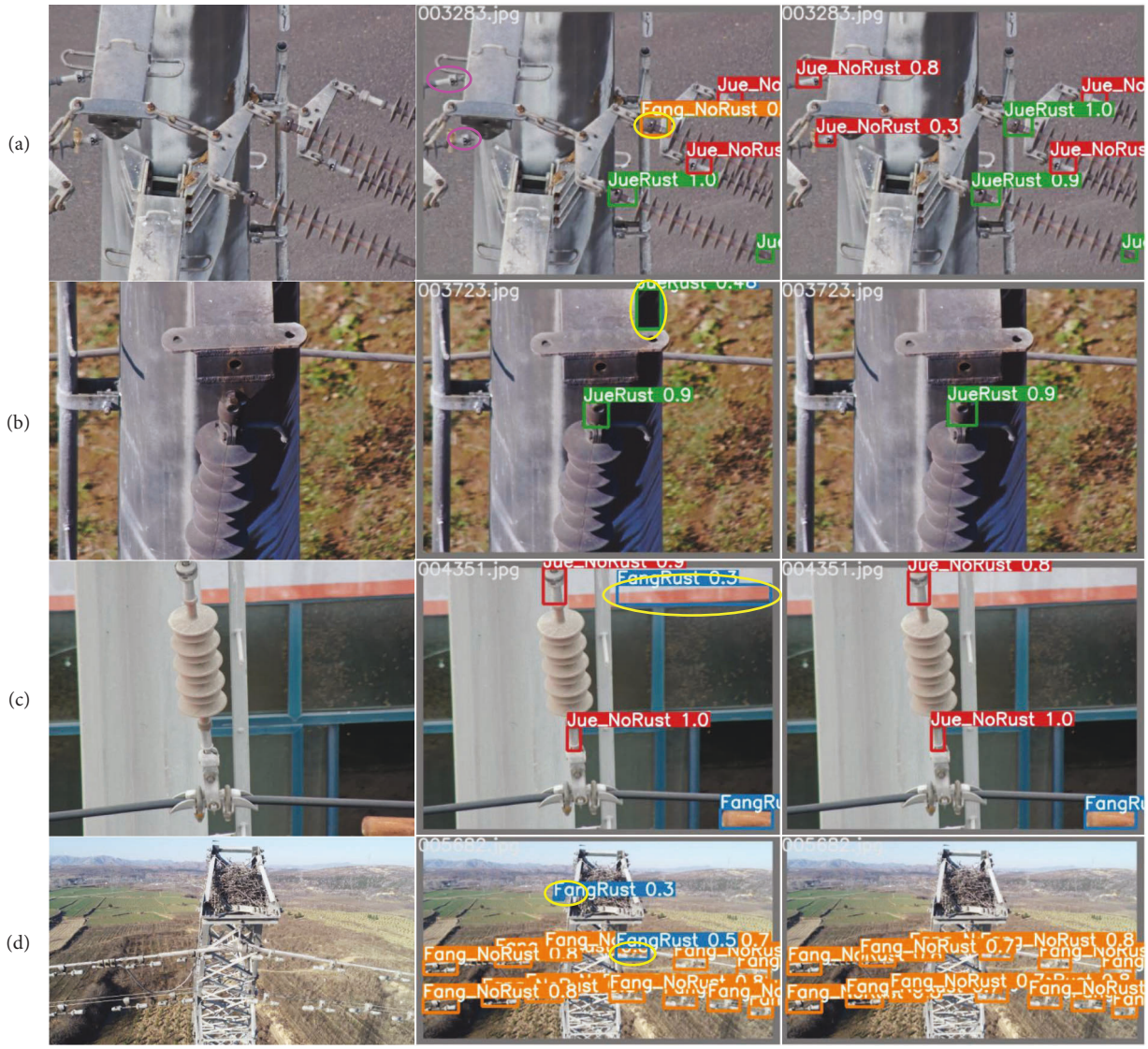

Figure 7: Detection results of the YOLOV5 and the PWR-YOLOV5 in real scenes.

middle. Meanwhile, the PWR-YOLOV5 detection outcomes are on the right. In the figure, the rectangular label boxes in blue, orange, green, and red, respectively, indicate FangRust, Fang_NoRust, JueRust, and Jue_NoRust, and the purple ellipse boxes denote missed detection, and the yellow ellipse boxes represent false detection. The YOLOV5 has missed detections of small objects in (a) and false detections in (a) and (d). In the case of serious background interference in (b), (c), and (d), some background items are mistakenly identified as targets. In contrast, the optimized PWR-YOLOV5 algorithm can significantly improve missed detections, false detections, and object identifications in complicated background.

3.3.3. Comparison with Other Classical Algorithms. For the purpose of validating the performance of the PWR-YOLOV5 algorithm, we chose representative one-stage algorithms SSD [23], RetinaNet [24], YOLOV3 [25], YOLOV4 [26], and two-stage algorithms Faster *R*-CNN [27], and on

Table 3: Performance comparison of different algorithms.

Models	AP (%)				mAP (%)	FPS	MS (MB)
	FangRust	Fang_NoRust	JueRust	Jue_NoRust			
SSD	83.35	91.84	78.45	51.78	76.36	44.7	96.61
RetinaNet	79.29	65.35	74.71	56.18	68.89	16.5	145.92
YOLOV3	75.43	93.83	72.50	45.12	71.72	31.4	93.83
YOLOV4	89.85	**98.42**	90.94	67.51	86.68	21.2	256.34
Faster *R*-CNN	83.54	92.17	66.54	47.22	72.37	22.4	547.02
YOLOX	90.33	96.38	92.60	71.21	87.63	25.5	36.03
CerterNet	85.94	94.41	85.45	63.76	82.39	37.8	131.01
YOLOV5	94.36	89.72	93.50	83.03	90.15	**109.9**	**14.40**
Ours	**96.88**	95.02	**95.61**	**93.96**	**95.37**	64.9	26.65

Table 4: Results of the ablation experiment.

Base	WA-PANet	PSA	RFEnet	EIoU loss	AP (%)				mAP/%	FPS	MS (MB)
					FangRust	Fang_NoRust	JueRust	Jue_NoRust			
✓					94.36	89.72	93.50	83.03	90.15	109.9	14.40
✓	✓				94.52	93.77	94.64	87.89	92.70	83.3	17.52
✓	✓	✓			94.80	93.80	97.90	89.33	93.96	68.5	22.09
✓	✓	✓	✓		95.79	93.93	97.30	93.83	95.21	64.9	26.65
✓	✓	✓	✓	✓	96.88	95.02	95.61	93.96	95.37	64.9	26.65

the basis of anchor-free algorithms CenterNet [28] and YOLOX [29], we developed contrast experiments. The test results of each model are shown in Table 3.

It can be seen from Table 3 that the detection accuracy of the proposed algorithm is optimal for the other three types of targets, except that the AP of Fang_NoRust is lower than YOLOV4 and YOLOX. Compared with the one-stage algorithm YOLOV4 with better detection performance, the AP of PWR-YOLOV5 on Fang_Rust, Jue_Rust, and Jue_NoRust is improved by 7.03%, 4.67%, and 26.45%, respectively. In comparison with the two-stage algorithm Faster *R*-CNN, the average precision of the improved algorithm for four classes was increased by 13.34%, 2.85%, 29.07%, and 46.74%, respectively. In contrast to YOLOX based on the anchor-free algorithm, the detection AP of the optimized YOLOV5 for Fang_Rust, Jue_Rust, and Jue_NoRust is enhanced by 6.55%, 3.01%, and 22.75%, respectively. Contrasted with the mAP, which measures the overall detection accuracy of the model, the PWR-YOLOV5 algorithm is the best, reaching 95.37%. In terms of detection speed and memory occupancy, the proposed algorithm is inferior to the YOLOV5, but better than other common detection models.

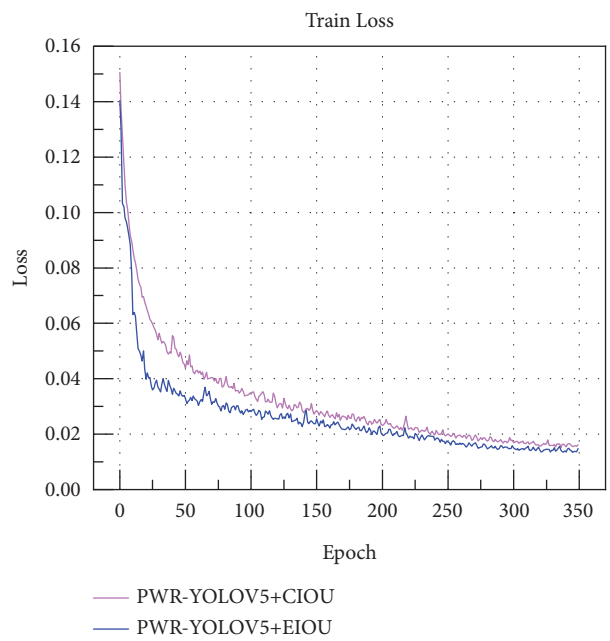

Figure 8: The training loss curves of CIOU Loss and EIOU Loss.

3.3.4. Ablation Experiment. In order to prove the influence of the four improved methods on the detection results, an ablation experiment is designed in this study, which chooses the YOLOV5 algorithm as the base. The experimental results are shown in Table 4.

All of the four proposed methods can improve the detection accuracy of various targets to a certain extent according to Table 4. The WA-PANet feature fusion network constructed by deepening the feature fusion process and introducing skip layer connections and adaptive feature fusion factors can not only ensure the integrity of fusion information but also take into account the influence of features from different stages on the fusion results. Compared with the YOLOV5 algorithm, the mAP advances by 2.55%. The PSA mechanism introduces channel attention on the feature maps of different receptive fields and adaptively associates spatial information and channel weights with $Softmax$ to produce pixel-level attention on objects, and the mAP is increased from 92.70% to 93.96%. The RFENet adopts bottleneck layers and dilated convolutions of different scales to strengthen the feature extraction for the

fusion output of the WA-PANet, which can make full use of features of different receptive fields and enhance the feature expression ability of various scale targets. The mAP of the algorithm advances to 95.21%. The three improvements referred to above to the model structure increase the number of parameters to different degrees, so the detection speed decreases somewhat, but it can still meet the requirement of real-time detection. In addition, the training loss curves of the improved algorithm under the CIOU Loss and the EIOU Loss are respectively drawn in Figure 8. By comparing the convergence of the two curves, it can be seen that the optimization method of the loss function proposed in this study makes the convergence rate faster and the positioning accuracy more superior.

4. Conclusions

Aiming at the problems of poor detection accuracy, serious missed detection, and false detection of corrosion components in distant view inspection by UAV, we propose a PWR-YOLOV5 detection method for corrosion components based on the YOLOV5 algorithm. This method firstly constructs a new feature fusion network, WA-PANet, which makes full use of the features at different stages and improves the detection accuracy of the corroded targets in remote view. Secondly, the PSA mechanism is led into the deep layers of the network to effectively suppress the background information and highlight the feature expression of the pixel-level targets. And then, the bottleneck structures and dilated convolutions of different scales are used to establish the feature enhancement network RFENet to alleviate the problem of feature expression ability weakening caused by feature fusion at different semantic levels. Finally, we apply the EIoU Loss to optimize the loss function of bounding box regression, so as to improve the positioning accuracy for the rusted targets.

The experimental results show that the proposed algorithm has more accurate location performance and higher detection accuracy. Compared with the original network, the average precision on FangRust, Fang_NoRust, JueRust, and Jue_NoRust are heightened by 2.52%, 5.3%, 2.11%, and 10.93%, respectively. The mAP can reach up to 95.37%, and the detection speed is 64.9FPS. Considering the accuracy and speed comprehensively, the proposed algorithm has higher application value, which provides a new idea for the corrosion component inspection of the overhead transmission line by UAV in the mining area.

Data Availability

The code data used to support the findings of the study can be obtained from the corresponding author upon request. But the image data involve in project privace, which is not publicly available.

Conflicts of Interest

The authors declare that the publication of this study does not refer to any conflicts of interest.

Acknowledgments

The authors are thankful to the staff of the Hemei Group Power Supply Department for their assistance in the process of this study. This study was funded by the Training Plan of Young Backbone Teachers in Colleges and Universities of Henan Province (Grant no. 2019GGJS060).

References

[1] L. Zhang, "Accelerating the transition of global clean energy Mineral security has become a national strategy," *China Economic Review*, vol. 2, no. 1, p. 8, 2021.

[2] J. Zhang, *Design and Implementation of Intelligent Remote Patrol Line System of Transmission Line*, University of Electronic Science and Technology of China, Chengdu, Sichuan, 2015.

[3] D. I. Jones, "An experimental power pick-up mechanism for an electrically driven UAV," in *Proceedings of the 2007 IEEE International Symposium on Industrial Electronics*, pp. 2033–2038, IEEE, Vigo, Spain, June 2007.

[4] O. Araar, N. Aouf, and J. L. Vallejo Dietz, "Power pylon detection and monocular depth estimation from inspection UAVs," *Industrial Robot: International Journal*, vol. 42, no. 3, pp. 200–213, 2015.

[5] S. Zhang, Z. Lei, and Y. Liu, "Research and analysis on intelligent inspection scheme of unmanned aerial vehicle based on artificial intelligence," *Chinese Journal of Sichuan Electric Power Technology*, vol. 42, no. 3, pp. 90–94, 2019.

[6] Z. Li, Y. Liu, R. Walker, R. Hayward, and J. Zhang, "Towards automatic power line detection for a UAV surveillance system using pulse coupled neural filter and an improved Hough transform," *Machine Vision and Applications*, vol. 21, no. 5, pp. 677–686, 2010.

[7] M. Recky and F. Leberl, "Windows detection using K-means in CIE-lab color space," in *Proceedings of the 2010 20th International Conference on Pattern Recognition*, pp. 356–359, IEEE, Istanbul, Turkey, August 2010.

[8] W. Song, D. Zuo, B. Deng, and H. Zhang, "Corrosion defect detection of earthquake hammer for high voltage transmission line," *Chinese Journal of Scientific Instrument*, vol. 37, no. 1, pp. 113–117, 2016.

[9] Y. Dai, D. Lv, and S. Guo, "Transmission line rusted area detection scheme based on color and texture features," *Chinese Journal of Industrial Control Computer*, vol. 31, no. 9, pp. 39-40, 2018.

[10] X. Huang, X. Zhang, Y. Zhang, and L. Zhao, "A method of identifying rust status of dampers based on image processing," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 8, pp. 5407–5417, 2020.

[11] H. K. Shen, P. H. Chen, and L. M. Chang, "Automated steel bridge coating rust defect recognition method based on color and texture feature," *Automation in Construction*, vol. 31, pp. 338–356, 2013.

[12] C. Sun, X. Qiu, X. Luo, Q. Huang, and C. Cao, "Rust defect recognition method for anti-vibration hammer based on the features of aggregated channels," *Chinese Journal of Computing Technology and Automation*, vol. 39, no. 2, pp. 128–132, 2020.

[13] L. Petricca, T. Moss, G. Figueroa, and S. Broen, "Corrosion detection using AI: a comparison of standard computer vision techniques and deep learning model," in *Proceedings of the Sixth International Conference on Computer Science,*

Engineering and Information Technology, pp. 91–99, IEEE, New Jersey, NJ, USA, May 2016.

[14] K. Zhang, Q. Hou, and W. Huang, "Defect detection of anti-vibration hammer based on improved faster R-CNN," in *Proceedings of the 2020 7th International Forum on Electrical Engineering and Automation IFEEA*, pp. 889–893, IEEE, Hefei, China, September 2020.

[15] Z. Wu, W. Xiong, J. Ren, and M. Jiang, "Corrosion object detection of power equipment based on lightweight SSD," *Chinese Journal of Computer Systems & Applications*, vol. 29, no. 2, pp. 262–267, 2020.

[16] H. Zhang, L. Wu, Y. Chen, R. Chen, and S. Kong, "Attention-Guided Multitask Convolutional Neural Network for Power Line Parts Detection," *IEEE Transaction on Instrumentation and Measurement*, vol. 71, Article ID 5008213, 2022.

[17] Yolov5, https://github.com/ultralytics/yolov5, 2021.

[18] M. Wu, Z. Zhang, Y. Song, Z. Shu, and B. Li, "Small object detection network based on adaptive feature enhancement," *Chinese Journal of Laser & Optoelectronics Progress*, vol. 59, pp. 1–14, 2022.

[19] H. Zhang, K. Zu, J. Lu, Y. Zou, and D. Meng, "EPSANet: An Efficient Pyramid Squeeze Attention Block on Convolutional Neural Network," 2021, https://arxiv.org/abs/2105.14447.

[20] J. Hu, S. Li, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR*, pp. 7132–7141, IEEE, Salt Lake City, Utah, USA, June 2018.

[21] S. Liu, D. Huang, and Y. Wang, "Receptive field block net for accurate and fast object detection," *Proceedings of the European Conference on Computer Vision ECCV September*, pp. 385–400, Munich, Germany, 2018.

[22] Y. F. Zhang, W. Ren, Z. Zhang, J. Zhen, W. Liang, and T. Tieniu, "Focal and Efficient IOU Loss for Accurate Bounding Box Regression," 2021, https://arxiv.org/abs/2101.08158.

[23] W. Liu, D. Anguelov, D. Erhan, E. Dumitru, S. Christian, and R. Scott, "SSD: single shot multibox detector," in *Proceedings of the European Conference on Computer Vision*, pp. 21–37, Springer International Publishing, Amsterdam, Netherlands, October 2016.

[24] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proceedings of the IEEE International Conference on Computer Vision ICCV*, pp. 2980–2988, IEEE, Venice Italy, October 2017.

[25] J. Redmon and A. Farhadi, "YOLOV3: An Incremental Improvement," 2018, https://arxiv.org/abs/1804.02767.

[26] 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.

[27] S. Ren, K. He, R. Girshick, J. Sun, and R.-C. N. N. Faster, "Towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[28] K. Duan, S. Bai, L. Xie, Q. Honggang, H. Qingming, and Q. Tian, "Centernet: keypoint triplets for object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision ICCV*, pp. 6569–6578, IEEE, Seoul, Korea, October 2019.

[29] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding yolo Series in 2021," 2021, https://arxiv.org/abs/2107.08430.