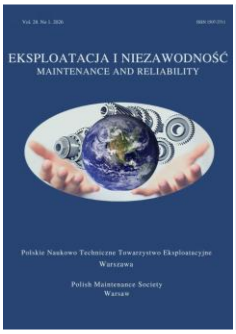


EksploatacjaiNiezawodnosc – Maintenance and Reliability
Volume 28 (2026), Issue 1
journal homepage: http://www.ein.org.pl

Article citation info:
Zhao J, Xiao Z, Bai B, Yang L, Li X, Defect Recognition of Transmission Line Unmanned Aerial Vehicle Inspection Images Based on Cascade R-CNN Algorithm, EksploatacjaiNiezawodnosc – Maintenance and Reliability 2026: 28(1)http://doi.org/10.17531/ein/207303

Defect Recognition of Transmission Line Unmanned Aerial Vehicle Inspection Images Based on Cascade R-CNN Algorithm

Indexed by:

Jie Zhao[a], Zhanhai Xiao[a], Bing Bai[a], Lei Yang[a], Xuanfeng Li[a,*]

[a] Xilingol Power Supply Branch of Inner Mongolia Electric Power (Group) Co., Ltd, China

Highlights

- ResNeXt152 backbone enhances multi-branch feature learning capability.
- Recursive feature pyramid integrates multi-scale defect information for accuracy.
- Focal Loss optimizes class imbalance in training samples.
- Image preprocessing improves defect feature expression in complex scenarios.
- Enhanced Cascade R-CNN achieves high-precision detection with low complexity.

Abstract

The representation and features of different defects in transmission lines in images vary greatly, resulting in incomplete feature extraction and affecting the accuracy of defect recognition. Therefore, a defect recognition method for transmission line unmanned aerial vehicle inspection images based on cascaded R-CNN algorithm is proposed. By preprocessing inspection images through denoising, enhancement, and normalization, ResNeXt152 is selected as the backbone network to improve the cascaded R-CNN algorithm, learning richer feature representations and focusing on different types of defect features; Introducing recursive feature pyramid structure for feature fusion and hierarchical prediction, capturing defect information at different scales; Utilizing focus loss to alleviate the problem of imbalanced positive and negative samples, in order to achieve defect recognition in transmission lines. The experimental results show that the proposed method has high defect recognition accuracy and low complexity, and can provide technical support for the safe operation of transmission lines.

Keywords
cascade R-CNN algorithm, transmission line, unmanned aerial vehicle inspection, image defect recognition, recursive feature pyramid, loss function

1. Introduction

As the power system evolves, transmission lines are becoming longer and more extensive in their reach, making their secure and stable operation indispensable for ensuring the grid's reliable power delivery [1]. However, traditional inspection methods, like manual checks, suffer from inefficiencies and accuracy constraints, rendering them inadequate for the present demands of transmission line inspection. Especially when facing the damage of key components such as insulators, shock absorbers, and pressure equalization rings, as well as the identification of attachments such as bird nests, the speed and accuracy of manual detection are greatly limited [2-4]. Therefore, the research on defect recognition methods for unmanned aerial vehicle inspection images of transmission lines is particularly important. UAV inspections enable swift and precise imaging of transmission lines and their vicinity, significantly enhancing inspection efficiency and precision. This is pivotal in maintaining the safe and stable operation of transmission lines while minimizing inspection expenses and hazards.

Current scholars have proposed various identification

(*) Corresponding author.
E-mail addresses: J. Zhao (ORCID: 0009-0004-5846-3662) xmdyjzhaojie@163.com, Z. Xiao (ORCID: 0009-0003-0200-5227) zhanhai1974@163.com, B. Bai (ORCID: 0009-0001-7902-9409) 18847952444@163.com, L. Yang (ORCID: 0009-0003-0232-5706) yl_sw@126.com, X. Li (ORCID: 0009-0006-6060-5644) 15352845658@163.com

methods, which not only improve the efficiency of defect identification, but also promote the further development and application of inspection technology. For example, Chen et al. [5]proposed a method for detecting tower defects in unmanned aerial vehicle inspection images based on an improved YOLOv4 model. The method replaces the CSPDarkNet53 network with a lightweight MobileNetV2 network, replaces max pooling with average pooling in the SPP module, and introduces attention mechanism CBAM to enhance feature representation. Although attention mechanism CBAM has been introduced to enhance feature expression, the implementation principle of CBAM is relatively complex, which may increase the computational burden of the model, especially on resource limited drone platforms. In addition, the effectiveness of CBAM largely depends on its parameter settings and training process, and improper parameter settings or insufficient training may limit its recognition performance. Wang et al. [6] proposes an intelligent defect recognition method for unmanned aerial vehicle inspection of transmission lines based on an offline Gaussian model. This method enhances the classifier in Mask R-CNN through an offline Gaussian model, which does not require additional training and is robust to data distribution. Although offline Gaussian models do not require additional training, the operation of the entire system still requires a certain amount of computational resources. In the real-time application scenario of drone inspection, if computing resources are limited, it may affect the recognition accuracy of the system. Ye et al. [7] proposes a defect detection method for unmanned aerial vehicle inspection of transmission line components based on an improved YOLOv3 model. The K-means++algorithm is introduced to solve the problem of insensitivity to small targets, the Focalloss function is introduced to solve the problem of sample imbalance, the Mish activation function is introduced to improve model accuracy, and the attention mechanism Senet is introduced to improve feature extraction performance. By comparing and analyzing the performance of the model before and after improvement, the superiority of this method has been verified. Although the introduction of various improvement mechanisms has improved the performance of the model, it may also lead to a more complex model structure, a significant increase in computational complexity, and an increase in FLOPs values. In actual drone inspection scenarios, higher FLOPs values mean that more computing resources and time are required to process image data, which may have high performance requirements for hardware devices. In addition, in situations where a large number of inspection images need to be processed in real-time, it may slow down the processing speed and fail to meet the real-time requirements, thereby limiting the promotion and use of this method in some practical application scenarios that have limitations on real-time and hardware resources.

Based on the above research, in order to improve the recognition effect, this paper studies a defect recognition method for unmanned aerial vehicle inspection images of transmission lines based on Cascade R-CNN algorithm, aiming to enhance the flexibility and accuracy of recognition while maintaining high efficiency, especially in complex environments and small target detection, demonstrating better performance.

2. Image preprocessing for unmanned aerial vehicle inspection of transmission lines

During the inspection process, images captured by drones may be contaminated by various types of noise, including Gaussian noise and salt-and-pepper noise, due to environmental conditions, equipment limitations, and other factors. Noise can affect the quality of images and reduce the accuracy of defect recognition. Therefore, it is necessary to denoise the images to provide reliable image data support for subsequent recognition. In this paper, the mean filtering technique [8] is employed to reduce noise in unmanned aerial vehicle (UAV) inspection images of transmission lines. Assuming that the neighborhood window size of a certain pixel point (x, y) in image $f(x, y)$ is $m \times n$, the pixel point after mean filtering is:

$$g(x, y) = \frac{1}{mn} \sum_{i=-\frac{m-1}{2}}^{\frac{m-1}{2}} \sum_{j=-\frac{n-1}{2}}^{\frac{n-1}{2}} f(x+i, y+j) \tag{1}$$

Overall, complete image denoising processing. Due to the possible issues of uneven lighting and low contrast in the inspection images of transmission lines, it is necessary to enhance the images to highlight the defect features in the images. This article uses histogram equalization method to enhance the denoised transmission line unmanned aerial vehicle inspection image. Assuming that the grayscale range of denoised image $f(x,y)$ is $[0,L-1]$ and its grayscale histogram is $h(r_k)(k=1,2,\dots,L-1)$, where r_k represents the k -th grayscale level and $h(r_k)$ represents the number of pixels with grayscale level r_k, the cumulative distribution function $CDF(r_k)$ of grayscale level r_k is:

$$CDF(r_k)=\sum_{i=0}^{k} h(r_i) \tag{2}$$

After histogram equalization, the new gray level s_k is:

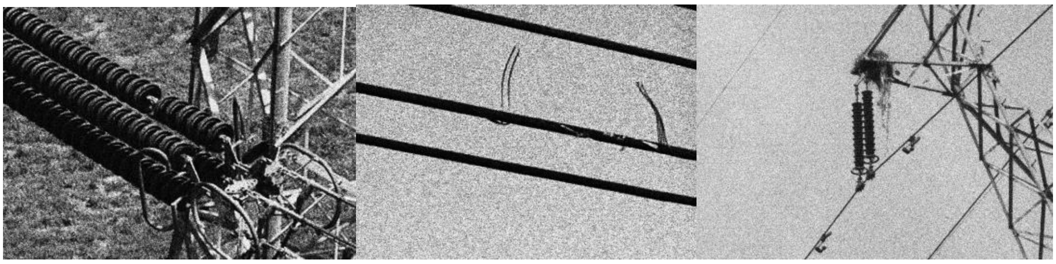
$$s_k=\frac{L-1}{MN}CDF(r_k) \tag{3}$$

Among them, M and N represent the number of rows and columns in the image.

Different drone equipment, shooting environments, and other factors may lead to differences in the grayscale range and size of inspection images, which can impact subsequent defect recognition algorithms. Therefore, it is necessary to normalize the image, by unifying the grayscale value range to a fixed interval and adjusting the image size to a fixed dimension. This study normalizes the drone inspection images of transmission lines using the grayscale normalization method. Assuming that the histogram equalization method enhances the grayscale value range of image $[a,b]$, which is $[c,d]$, and the normalized grayscale value range is $w(x,y)$, the normalized image is:

$$w(x,y)=\frac{d-c}{b-a}(f'(x,y)-a)+c \tag{4}$$

Select a representative set of drone inspection raw images of transmission lines as the test dataset to verify the effectiveness of the preprocessing method. The results are shown in Figure 1.



(a) Before pre-processing

(b) After pre-processing

Figure 1. Preprocessing effect of unmanned aerial vehicle inspection images for transmission lines.

As evident from Figure 1, the preprocessed image has exhibited notable improvements in terms of noise reduction, effectively suppressing noise in the image and making it smoother and clearer. Preprocessing the drone-captured transmission line inspection images has significantly enhanced their quality, making defect features more prominent and establishing a solid foundation for subsequent defect recognition tasks in these images.

During the acquisition process of drone inspection images for transmission lines, there are problems such as noise, uneven lighting, low contrast, and differences in grayscale value range and size due to environmental and equipment factors. To address these issues, this article adopts mean filtering for denoising, enhances the image through histogram equalization to highlight defect features, and uses grayscale normalization method to unify the grayscale range and adjust the size. These preprocessing operations help improve image quality and lay the foundation for defect recognition in subsequent drone inspection images of transmission lines. The following will focus on exploring the relevant content of defect recognition in drone inspection images of transmission lines.

3. Design of Defect Recognition Method for Unmanned Aerial Vehicle Inspection Images of Transmission Lines

In the drone inspection images of transmission lines, there may be significant differences in the size, shape, and characteristics of defects. The Cascade R-CNN algorithm is based on deep convolutional neural networks and can automatically learn complex features in transmission line inspection images. By stacking multiple convolutional layers, the network can gradually extract features from low to high levels, fuse features from different levels, provide detailed information about defects, and ensure the accuracy and reliability of defect recognition. Therefore, using the processed drone inspection image of the transmission line as input, the Cascade R-CNN algorithm is used to complete defect recognition.

3.1. Cascade R-CNN Algorithm Principle

The Cascade R-CNN network structure is a multi-stage extension based on the Faster R-CNN algorithm, as shown in Figure 2. 'In' is the input target image; Conv "is a convolutional neural network used for feature extraction; Pool "is the pooling process for feature extraction; 'H' is the network header; B "is the boundary regression box; C "is the classification. Both types of networks perform object detection in two steps: first, target localization, and then target classification.

The specific steps of the Cascade R-CNN network are as follows: firstly, using convolutional neural networks Res Net or VGG16 as the backbone network for feature extraction, the input image is subjected to feature extraction to generate a feature map; Secondly, the image undergoes the identification of potential target areas via a Region Proposal Network (RPN). Subsequently, the Region of Interest (ROI) network gathers all proposed bounding boxes, computes their respective feature maps, and forwards these to the subsequent network stage. Ultimately, the samples progress through a series of cascaded stages where they undergo regression and classification training, leveraging a combination of cascaded bounding box refinement and cascaded detection techniques.

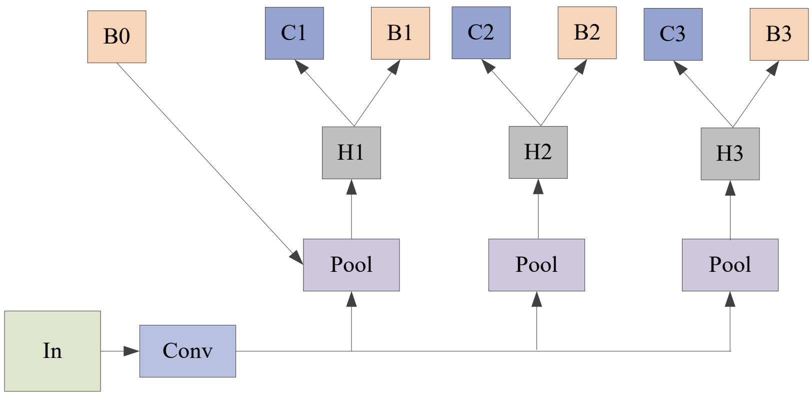


Figure 2. Schematic diagram of Cascade R-CNN network structure.

3.2. Cascade R-CNN Algorithm Improvement

Based on the characteristics of unmanned aerial vehicle inspection images for transmission lines, this paper has made three improvements on the Cascade R-CNN network to better extract and fuse complex and diverse defect features, and improve the algorithm's ability to detect and identify various defects. Firstly, the backbone network chose ResNeXt152, which is an upgraded version of ResNet that introduces the idea of multi branch Inception; Compared to ResNet, with the same number of parameters, the results are better. Secondly, for unmanned aerial vehicle inspection images of transmission lines with different scales [9-10], a recursive feature pyramid (RFP) structure is introduced for feature fusion and hierarchical prediction of targets at different scales. The feature maps of the backbone network are processed through an atrous space search (ASS) module for recursion and fusion, improving the accuracy of defect recognition. Finally, in order to address the issue of an increase in the number of negative samples and an imbalance in the number of positive and negative samples caused by the generation of a certain number of anchor boxes centered around pixels in feature maps of different scales during RPN training, which is much larger than the number of defect targets in unmanned aerial vehicle inspection images of transmission lines, Focal loss is used to improve the overall loss function and alleviate the impact of a smaller number of positive samples than negative samples on recognition accuracy.

3.2.1. ResNeXt152

The drone inspection images of transmission lines have specific defect characteristics. In order to more effectively detect and identify these defects, it is necessary to improve the existing object detection network [11-12]. Cascade R-CNN is an effective object detection network, but there is still room for improvement when facing special situations in transmission line inspection images. By replacing the backbone network of the Cascade R-CNN network and selecting ResNeXt152 as the new backbone network, it is expected to utilize its structural characteristics to improve the network's detection performance for inspection image defects.

The fundamental concept behind ResNet involves incorporating residual blocks as a means of addressing the issues of vanishing and exploding gradients in deep neural networks. The input signal in the residual block can skip some layers and add the outputs of subsequent layers directly, making it easier for the network to learn identity mapping. A typical residual block can be represented as:

$$y = F(x, \{W_i\}) + x \tag{5}$$

Among them, x is the input of the residual block, which is the preprocessed transmission line unmanned aerial vehicle inspection image, $F(x, \{W_i\})$ is the function that needs to be learned in the residual block, W_i represents the weight parameters in the function, and y is the output of the residual block.

ResNeXt introduces the multi branch idea of Inception based on ResNet. The Inception architecture is capable of capturing a more diverse set of features by applying convolution operations across multiple scales. ResNeXt incorporates the idea of multiple branches into residual blocks, where each residual block contains multiple parallel paths with the same topology but different weight parameters. Assuming a ResNeXt residual block has C branches, each with an output of $F_i\left(x, \{W_{i,j}\}\right)$, based on formula (5), the output of the residual block can be improved and expressed as:

$$y = \sum_{i=1}^{C} F_i\left(x, \{W_{i,j}\}\right) + x \tag{6}$$

The features obtained from ResNeXt residual blocks are fed into a Region Proposal Network (RPN) to produce potential regions that likely contain defects.

By selecting ResNeXt152 as the backbone network based on the Cascade R-CNN network and utilizing ResNeXt's multi branch idea, the detection performance of the network for defects in unmanned aerial vehicle inspection images of transmission lines can be improved without increasing the number of parameters. This improvement provides more effective technical support for intelligent inspection of transmission lines, helping to timely detect and handle line defects, and ensuring the safe and stable operation of the power system.

3.2.2. Recursive Feature Pyramid

This article adopts a recursive feature pyramid (RFP) structure for feature fusion and hierarchical prediction of targets at different scales. The feature maps of the backbone network are processed through an atrous space search (ASS) module for recursion and fusion.

Feature Pyramid Networks (FPN) is an object detection approach based on deep convolutional neural networks that constructs a hierarchical, multi-scale feature pyramid by organizing deep convolutional layers into a pyramid-like structure. This article adopts the method of assigning defects in unmanned aerial vehicle inspection images of different scales to different feature layers for prediction, to solve the problem of inaccurate prediction caused by severe feature loss in unmanned aerial vehicle inspection images of transmission lines [13-15]. The recursive feature pyramid leverages the pyramid structure, repeatedly harnessing the feature extraction capabilities of the backbone network to progressively integrate and generate feature maps in an iterative manner. In contrast to the traditional feature pyramid, the recursive feature pyramid employs a hierarchical approach for predicting defects in transmission line images captured by unmanned aerial vehicles. Additionally, it enhances the feature representation of these defects in each feature map through the processes of feature regression and fusion.

Assuming K_i represents the i-stage convolution process of the bottom-up backbone network, R_i represents the i-stage feature map fusion process of the feature pyramid network, the feature map predicted by the target of the feature pyramid network is $f_i(i = 1,2,...,D)$, and the feature map after the convolution operation of the backbone network is $y_i(i = 1,2,...,D)$, where D represents the number of feature layers in the feature map, the output feature map is represented as follows:

$$\begin{cases} f_i = R_i(f_{i+1}, y_i) \\ y_i = K_i(y_{i-1}) \end{cases} \tag{7}$$

In contrast to Feature Pyramid Networks (FPN), the Recursive Feature Pyramid (RFP) introduces feedback connections, where the input is bifurcated into two streams: one serving as feedback input to the FPN (left branch) and the other originating from the original image (bottom-up input). T_i represents the preprocessing before feature fusion of the i -stage feedback input feature map, G_i is defined as the output feature of RFP, and v_i represents the i -stage feature map output by the backbone network during the first recursion.

Recursive structures can be used for multiple iteration processes, and the formula for the t -th recursive iteration process is as follows:

$$\begin{cases} G_i^t = R_i^t(G_{i+1}^t, v_i^t) \\ v_i^t = K_i^t\left(f_i^t, T_i^t(G_i^{t-1})\right) \end{cases} \tag{8}$$

Among them, $t = 1,2,...,T$ represents the number of recursive iterations of pyramid features.

3.2.3. loss function

During the RPN training process, a certain number of anchor boxes will be generated around the pixels of feature maps of different scales, which is much larger than the number of defect targets in the inspection images of transmission line drones, resulting in an increase in the number of negative samples. In addition, for small scale targets in images, the proportion of background exceeding that of small targets can also cause an imbalance in the number of positive and negative samples. Therefore, this article adopts Focal loss to improve the overall loss function and alleviate the impact of having fewer positive samples than negative samples on recognition accuracy. The overall loss function of the improved Cascade R-CNN is as follows:

$$L(o_i, o_i', q_i, q_i') = L_{Focal}(o_i, o_i') + L_{cls}(o_i, o_i') + \varphi L_{loc}(q_i, q_i') \tag{9}$$

$$L_{Focal} = \begin{cases} -\delta(1 - o_i)^{\varphi} \log_2 o_i \; o_i' = 1 \\ -(1 - \delta)(o_i)^{\varphi} \log_2 o_i \; o_i' = 0 \end{cases} \tag{10}$$

$$L_{loc}(o_i, o_i') = Y(o_i - o_i') \tag{11}$$

$$o_i = \frac{e^{h_j}}{\sum_{k=1}^{m} e^{h_k}} \tag{12}$$

$$o_i' = \begin{cases} 1, IoU(x, g) > u \\ 0, other \end{cases} \tag{13}$$

Among them, L_{Focal} represents Focal loss; L_{cls} represents classification loss L_{loc}; Indicating positioning loss; o_i signifies the likelihood that the anchor box corresponds to a positive sample; o_i' represents the category of the anchor box; q_i represents the anchor box parameter vector; q_i' represents the marker box parameter vector; Y represents the smooth function; φ represents the trade-off coefficient; Y represents the smooth function; φ represents the trade-off coefficient; h_j represents the j -th value of the Softmax output vector, $j = 1,2,...,m$, m represent the number of categories; x represents anchor box; g represents the real bounding box; u represents the IoU threshold; $IoU(x, g)$ denotes the ratio of intersection to union between the anchor box and the actual bounding box.

Based on the improved overall loss function, classify and judge each candidate region to determine whether it belongs to a certain defect category of the transmission line. Once the defect area is identified for annotation, the annotation process involves drawing an annotation box on the original image to clearly identify the defect area. When drawing the annotation box, the center coordinates x and y of the border, as well as the width w and height h of the box, are normalized. The expression is as follows:

$$\partial_x = \frac{g_x - q_x}{q_w} \tag{14}$$

$$\partial_y = \frac{g_y - q_y}{q_h} \tag{15}$$

$$\partial_w = \log_2(g_w / q_w) \tag{16}$$

$$\partial_h = \log_2(g_h / q_h) \tag{17}$$

In the formula, g_x represents the position of the center coordinate x of the real bounding box; q_x represents the anchor box parameter vector under the center seat x marker; g_y is the center coordinate y position of the real bounding box; Anchor box parameter vector with q_y as the center coordinate y; q_w is the width of the anchor frame; q_h is the height of the anchor frame; g_w is the actual bounding box width; g_h is the actual bounding box height.

This paper proposes three improvements to the Cascade R-CNN algorithm based on the characteristics of unmanned aerial vehicle inspection images for transmission lines. Firstly, the backbone network is replaced with ResNeXt152, which introduces the multi branch idea of Inception based on ResNet to improve detection performance without increasing the number of parameters; Secondly, the Recursive Feature Pyramid (RFP) architecture is incorporated to facilitate feature fusion and enable hierarchical predictions for targets of varying scales. After processing the backbone network feature map through the hollow space search module, recursion and fusion are performed, which can optimize feature information compared to classical feature pyramids; Thirdly, to address the issues of increasing negative sample count and imbalanced positive and negative sample count in RPN training, Focal loss is adopted to improve the overall loss function and alleviate its impact on recognition accuracy.

3.3. Implementation of drone inspection image defect recognition for transmission lines

In section 3.2, the enhanced Cascade R-CNN algorithm was employed to detect defects in unmanned aerial vehicle (UAV) inspection imagery of transmission lines, ultimately enhancing the precision of defect recognition. The workflow of this algorithm is depicted in Figure 3, with its key procedures outlined below:

(1) Preprocess the drone inspection images of transmission lines. Firstly, due to environmental and equipment reasons, the images may contain Gaussian noise, salt and pepper noise, etc. The mean filtering method can effectively improve image quality and reduce the impact of noise on defect recognition accuracy by setting the pixel neighborhood window size to denoise the images; Secondly, to address the potential issues of uneven lighting and low contrast in inspection images, the histogram equalization method enhances the image by calculating the cumulative distribution function of gray levels and other operations, highlighting the defect features in the image; Finally, considering that different factors can affect the defect recognition algorithm due to differences in the grayscale value range and size of inspection images, the grayscale normalization method unifies the grayscale value range of the image to a fixed interval and adjusts the size to a fixed size. By preprocessing the image dataset, standardized image data can be provided for subsequent defect recognition algorithms, which helps improve the accuracy and reliability of defect recognition.

(2) Based on the preprocessed image dataset, construct a sample set for defect target recognition in unmanned aerial vehicle inspection images of transmission lines. 80% of the same batch in the constructed sample set is used as the training set, 20% is used as the validation set, and the line images in open scenarios are used as the validation set for training, validation, and testing the model.

(3) Input the preprocessed image dataset and use the Convolutional Neural Network ResNeXt152 as the backbone network for feature extraction to generate feature maps. As mentioned earlier, ResNeXt152 introduces the multi branch idea of Inception based on ResNet. Each residual block contains multiple parallel branches, and its residual block output is obtained by adding the outputs of each branch. The output of the residual block is obtained through equation (6). Transfer the features extracted from ResNeXt residual blocks to a Region Proposal Network (RPN) to generate candidate regions that may contain defects.

(4) Once the Region Proposal Network (RPN) identifies potential target areas in the image, the Region of Interest (ROI) network then gathers all proposed bounding boxes, computes their feature maps, and forwards these maps to the subsequent network stages.

(5) Using recursive feature pyramid (RFP) structure for feature fusion and hierarchical prediction of targets at different scales, obtained through equation (7); Then, the feature map of the backbone network is processed through the Empty Space Search (ASS) module and recursively fused to output the feature map, which can be obtained through equation (8).

(6) Using the training set to iteratively train the improved Cascade R-CNN model, the network parameters are adjusted through backpropagation algorithm to continuously learn the defect features in the drone inspection images of transmission lines, namely x, f_i, y_i, G_i.

(7) Input the target image to be recognized into the trained model, determine whether there are defects in the unmanned aerial vehicle inspection image of the transmission line, and annotate the final defect area to be recognized. That is, draw a annotation box on the original image and normalize the

boundary box parameters. The normalization process is implemented through equations (14) to (17), and output the image results with defect annotations.

(8) Finally, save the image with defect annotations as the recognition result output.

Through the above steps, the trained and improved Cascade R-CNN model was used to achieve defect recognition of unmanned aerial vehicle inspection images of transmission lines, and output image results with defect annotations.

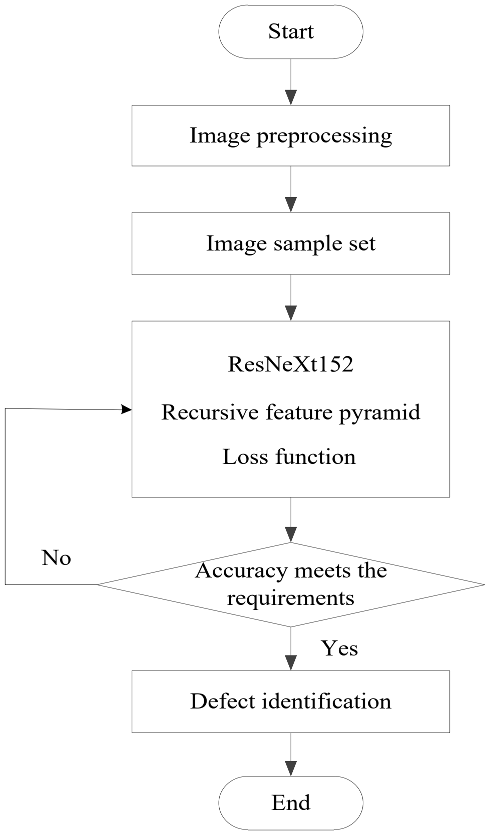


Figure 3. Flow chart of defect recognition in unmanned aerial vehicle inspection images of transmission lines.

4. Experimental verification

4.1. Dataset Establishment and Processing

The original images used in this experiment are 2500 drone aerial inspection samples taken by a certain power supply bureau in recent years. To expand the sample, this article adds Gaussian noise, multiplicative noise, and salt and pepper noise to the dataset images. Brightness transformation, scaling, image quality adjustment, enhancement filtering, and motion blur processing are used to simulate the real shooting environment of unmanned aerial vehicles, reflecting the authenticity of images taken with different resolutions, weather, lighting, time periods, and drone shaking.

After the above processing, the dataset images were finally expanded to 5000 images. Split the training set and testing set randomly in a proportion of 80% to 20%, ensuring that they remain mutually exclusive throughout the experimental process. The training set includes 4000 images, including 1203 insulators, 598 insulator defects, 405 shock absorbers, 794 phase rods, and 1000 bird nests. The test set includes 1000 images, including 387 insulators, 256 insulator defects, 165 shock absorbers, 102 phase rods, and 90 bird nests. As shown in Table 1.

Table 1. Experimental Dataset.

Type	Training set	Test set
Insulator	1203	387
Defects in insulators	598	256
Shockproof hammer	405	165
Alternating rod	794	102
Bird's nest	1000	90

The example image is shown in Figure 4.

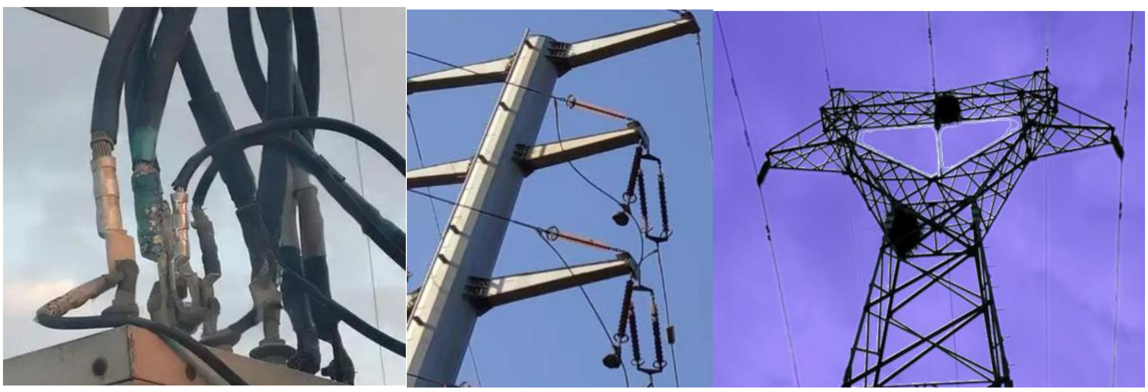


Figure 4. Sample Image Example.

4.2. Experimental setup

Based on the above test samples, the hardware environment, software tools, model parameters, and other settings related to the experiment are shown in Tables 2, 3, and 4, respectively.

Table 2. Hardware Environment.

Hardware components	Configuration parameter
CPU	Intel Core i9-10900K @ 3.7GHz
GPU	NVIDIA RTX 3090 (24GB VRAM)
Memory	64GB DDR4 @ 3200MHz
Storage	1TB NVMe SSD + 4TB HDD
Operating system	Ubuntu 20.04 LTS

Table 3. Experimental Software Tool Configuration

Software tool	Configuration
Programming language	Python 3.8
Deep Learning Framework	PyTorch 1.9.0
Image processing library	OpenCV 4.5.2
Data augmentation tools	NVIDIA Apex
Visual Tools	TensorBoard 2.6.0

Table 4. Model Parameter Settings.

Parameter	Numerical/Configuration
Network	ResNeXt152
Enter image size	1024×1024
Recursive iteration times	3
Focal Loss parameter	2.0
Focal Loss parameter	2.5
Classification loss weight	1.0
Positioning loss weight	0.5
Batch size	8
Initial learning rate	0.001
Training epochs	100
Non maximal inhibition threshold	0.5
Confidence threshold	0.6

4.3. Evaluation indicators

This article evaluates the performance of the algorithm using the standard metrics of average precision (AP) and mean average precision (mAP) across all categories in object recognition algorithms. The respective formulas for these calculations are outlined below:

$$AP = \frac{\sum_n Precision}{n} \tag{18}$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{19}$$

Among them, n is the total number of samples; N is the number of categories. Using recall rate as the horizontal axis and precision rate as the vertical axis, a curve can be obtained, and the area under this curve is the AP value.

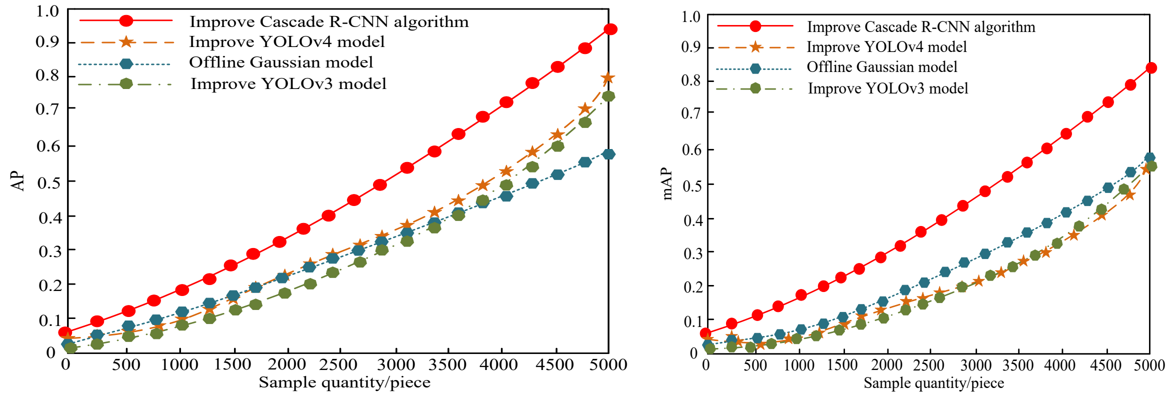
In addition, floating point operands (FLOPs) are selected as evaluation indicators: FLOPs are an important evaluation indicator used to evaluate target recognition algorithms, and their values represent the complexity of the target recognition algorithm. The larger the value, the higher the complexity of the algorithm. Its calculation formula is:

$$FLOPs = HW(C_{in} \times K)C_{out} \tag{20}$$

Among them, H and W stand for the height and width of the resultant feature matrix, respectively; C_{in} denotes the depth of the input feature matrix; C_{out} signifies the depth of the output feature matrix; and K represents the dimension of the convolutional kernel.

4.4. Analysis of Recognition Effect

To compare the effectiveness of the improved Cascade R-CNN algorithm with other methods, under the same dataset partitioning benchmark, the improved YOLOv4 model, offline Gaussian model, improved YOLOv3 model, and the improved Cascade R-CNN algorithm were selected for comparative experiments. The comparison results of AP and mAP values of different methods are shown in Figure 5.



(a) AP value(b) MAP value

Figure 5. Comparison results of defect recognition accuracy.

From Figure 5, it can be seen that in the actual scenario of unmanned aerial vehicle inspection of transmission lines, image data often has complexity and diversity, with various lighting conditions, background interference, and defects of different scales and forms. The improved YOLOv4 model, offline Gaussian model, and YOLOv3 model have exposed their limitations in feature extraction, target localization, and classification recognition when facing these complex situations, resulting in their inability to effectively capture defect features in unmanned aerial vehicle inspection images of transmission lines, and thus difficult to accurately identify various types of defects. In sharp contrast, the improved Cascade R-CNN algorithm demonstrated excellent performance, with an AP value of 0.958 and an mAP of 0.823. This algorithm adopts a series of improvement measures, such as optimizing the selection of the backbone network, introducing recursive feature pyramid structure for feature fusion, and using Focal loss to improve the overall loss function, to enhance its feature learning ability and adaptability to complex scenes. Therefore, the improved Cascade R-CNN algorithm can effectively distinguish multiple types of defects in transmission lines, demonstrating significant advantages in defect recognition tasks in unmanned aerial vehicle inspection images of transmission lines, and outperforming other comparative methods.

Utilizing formula (20), the floating-point operations per second (FLOPs) for various methods were computed, with the outcomes presented in Figure 6.

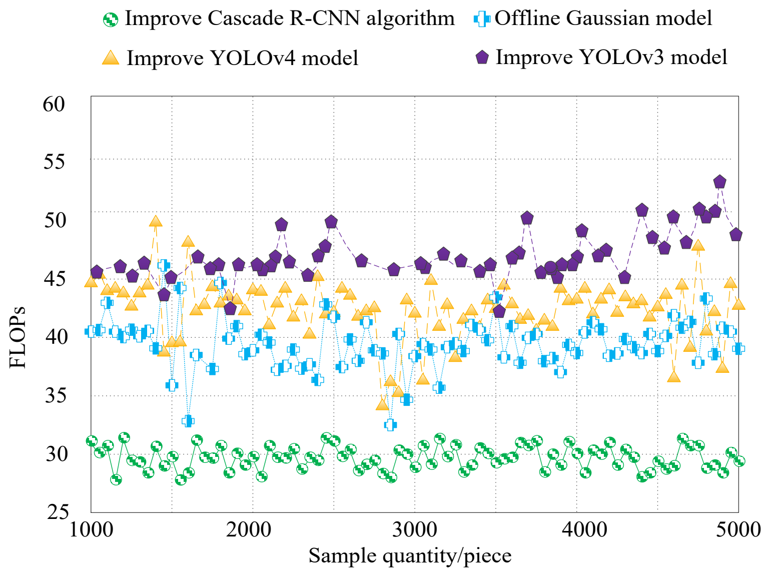


Figure 6. FLOPs Comparison Results.

In practical application scenarios of unmanned aerial vehicle (UAV) inspections of transmission lines, computational resources are frequently constrained, particularly when handling vast amounts of inspection image data in real-time. Consequently, the computational efficiency of the model is of utmost importance. Observing Figure 6, it is evident that the FLOPs values of the enhanced YOLOv4 model, offline Gaussian model, and improved YOLOv3 model are relatively high, indicating that they require a large number of floating-point operations when performing defect recognition in unmanned aerial vehicle inspection images of transmission lines. This not only puts higher demands on the computing power of hardware devices, but may also lead to slower processing speed, which cannot meet the real-time requirements. The improved Cascade R-CNN algorithm can effectively reduce FLOPs values while maintaining a high defect recognition accuracy, always around 30. This is due to its improvements in network architecture design, feature extraction and fusion methods, as well as loss function optimization. By reasonably adjusting and optimizing the various components of the model, the improved Cascade R-CNN algorithm achieves a good balance between computational efficiency and recognition performance, making it more valuable for practical unmanned aerial vehicle inspection of transmission lines.

Finally, Figure 7 visually demonstrates the defect recognition performance of the improved Cascade R-CNN algorithm.

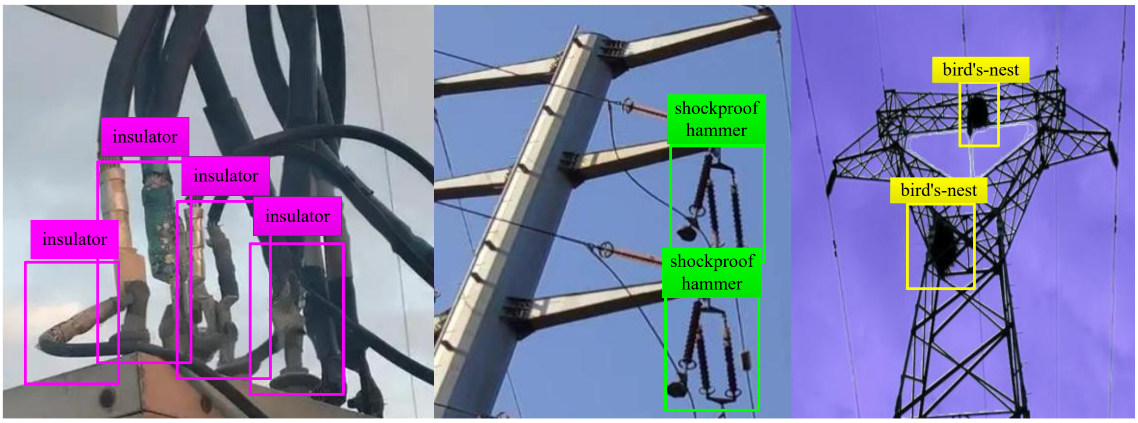


Figure 7. Defect recognition effect.

Observing Figure 7, it becomes apparent that the refined Cascade R-CNN algorithm is capable of accurately detecting defects in transmission line sample images. With advanced network structure and optimized algorithm mechanism, it can dynamically adjust the weight of feature fusion based on the importance of different levels of feature maps. When processing transmission line images, algorithms can better integrate feature information of different scales, thereby accurately capturing complex and diverse defect features. This powerful recognition capability provides reliable technical support for the safety inspection and maintenance of transmission lines, enabling efficient image acquisition through devices such as drones or intelligent cameras, and quickly and accurately identifying potential defects. It provides accurate location and type information for the maintenance and repair of transmission lines, ensuring their stable operation and reducing the risk of power outages caused by line failures.

5. Conclusion

The identification of defects in drone inspection images of transmission lines faces many challenges, such as the diversity of defect targets, scale variability, and imbalance of positive and negative samples. The defect recognition method based on Cascade R-CNN algorithm proposed in this article involves preprocessing of inspection images and various improvements to the algorithm. The experimental results show that the proposed method effectively improves the accuracy of defect recognition, maintains high recognition performance in complex backgrounds, reduces complexity, and achieves good recognition results. It provides reliable technical support for the safe operation of transmission lines and has important application value and practical significance. The specific manifestations are as follows:

1. Superior recognition accuracy:

The improved Cascade R-CNN algorithm has demonstrated excellent performance, with an AP value of 0.958 and mAP of 0.823, significantly higher than the comparison method. It can effectively achieve target defect detection, indicating that the recursive feature pyramid structure can effectively fuse multi-scale features and enhance sensitivity to targets.

2. Optimization of computational efficiency:

By using the ResNeXt152 backbone network and Focal Loss, the model can effectively reduce FLOPs values while maintaining high accuracy, always around 30, which can effectively meet the real-time requirements of drone inspection.

In summary, the method proposed in this article has been improved in multiple dimensions and outperforms existing technologies in terms of accuracy and efficiency, providing reliable technical support for intelligent inspection of transmission lines.

Although the defect recognition method for unmanned aerial vehicle inspection images of transmission lines based on Cascade R-CNN algorithm proposed in this article has achieved good results, there is still room for further improvement and expansion. In the future, we will explore the combination of this method with other advanced technologies, such as artificial intelligence reinforcement learning, edge computing, etc., to achieve a more intelligent and real-time defect identification and early warning system, provide a stronger technical guarantee for the safe and stable operation of transmission lines, and help the efficient development of the power industry.

References

1. Wang H, Yang Q, Zhang B,Gao D. Deep learning based insulator fault detection algorithm for power transmission lines. Journal of Real-Time Image Processing, 2024, 21(4):1-16. https://doi.org/10.1007/s11554-024-01495-9
2. Khalili M, Namdari F, Rokrok E. A fault location and detection technique for STATCOM compensated transmission lines using game theory. IET Generation, Transmission & Distribution, 2021,36(02):47-55. https://doi.org/10.1049/gtd2.12126
3. Liu Y, Wang B, Zheng X, Lu D, Fu M, Tai N. Fault location algorithm for non-homogeneous transmission lines considering line asymmetry. IEEE Transactions on Power Delivery, 2020, 36(9):1-10. https://doi.org/10.1109/TPWRD.2020.2968191
4. Shakiba F M, Azizi S M, Zhou M, Abusorrah A. Application of machine learning methods in fault detection and classification of power transmission lines: a survey. Artificial Intelligence Review, 2023, 56(7):5799-5836. https://doi.org/10.1007/s10462-022-10296-0
5. Chen J, An Z, Tang Z, Lu Z. Research on tower defect detection method of UAV patrol inspection image based on improved YOLOv4 model. Electrical Measurement & Instrumentation, 2023,60(10):155-160.
6. Wang H, Chen Y, Zhang X, Wu M, Mao W, Li H. Research on intelligent recognition method of transmission line UAV inspection defects based on offline Gaussian model. Electrical Measurement & Instrumentation, 2022,59(3):92-99.
7. Ye X, Sun J, Gan Y, Ran Q, Wu D, Lv Z. Application of improved YOLOv3 model in defect detection of transmission line components in UAV patrol inspection. Electrical Measurement & Instrumentation, 2023,60(5):85-91.
8. Geng H, Xu Y, Wu W. Aliasing Suppression Method for a Three-Phase Grid-Connected Photovoltaic Inverter Based on Multi-Sampling and Mean Filtering. Energies (19961073), 2024, 17(4):1-15. https://doi.org/10.3390/en17040907
9. Ahmed. F, Mohanta J C, Sanyal A. Inspection and identification of transmission line insulator breakdown based on deep learning using aerial images. Electric Power Systems Research, 2022, 211(10):1-15. https://doi.org/10.1016/j.epsr.2022.108199
10. Verrax P, Alglave N, Bertinato A, Kieffer M, Raison B. Low-complexity graph-based traveling wave models for HVDC grids with hybrid transmission lines: Application to fault identification. Electric Power Systems Research, 2022, 205(4):107710.1-107710.15. https://doi.org/10.1016/j.epsr.2021.107710
11. Wang D, Liu H, Zhang D, Qiao F, Liu X, Hou M. Travelling wave fault location for LCC-MMC-MTDC transmission lines based on frequency-dependent characteristic. International Journal of Electrical Power and Energy Systems, 2022, 138(6):107863.1-107863.8. https://doi.org/10.1016/j.ijepes.2021.107863
12. Shu H, Wang G, Tian X, Ju X, An N, Bo Z. MMC-HVDC line fault identification scheme based on single-ended transient voltage information entropy. International Journal of Electrical Power and Energy Systems, 2022, 141(10):107817.1-107817.14. https://doi.org/10.1016/j.ijepes.2021.107817
13. Qu J, Yin J, Jiang Y, Huang W, Chen Q. A sub-pixel convolution-based improved bidirectional feature pyramid network for pansharpening. Remote Sensing Letters, 2023, 14(3):91-101. https://doi.org/10.1080/2150704X.2022.2162350
14. Huangfu P, Dang L. A multi-scale pyramid feature fusion-based object detection method for remote sensing images. International Journal of Remote Sensing, 2023, 44(23/24):7790-7807. https://doi.org/10.1080/01431161.2023.2288947
15. Chen X, Li Y, Nakatoh Y. Pyramid attention object detection network with multi-scale feature fusion. Computers and Electrical Engineering, 2022, 104(Pt.1):108436-108446. https://doi.org/10.1016/j.compeleceng.2022.108436