Retraction

Retracted: A High-Efficiency Deep-Learning-Based Antivibration Hammer Defect Detection Model for Energy-Efficient Transmission Line Inspection Systems

International Journal of Antennas and Propagation

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] F. Zhou, G. Wen, G. Qian et al., "A High-Efficiency Deep-Learning-Based Antivibration Hammer Defect Detection Model for Energy-Efficient Transmission Line Inspection Systems," *International Journal of Antennas and Propagation*, vol. 2022, Article ID 3867581, 12 pages, 2022.

Research Article

A High-Efficiency Deep-Learning-Based Antivibration Hammer Defect Detection Model for Energy-Efficient Transmission Line Inspection Systems

Fangrong Zhou,[1] **Gang Wen,**[1] **Guochao Qian,**[1] **Yutang Ma,**[1] **Hao Pan,**[1] **Jing Liu,**[2] **and Jiaying Li** [3]

[1]*Joint Laboratory of Power Remote Sensing Technology (Electric Power Research Institute, Yunnan Power Grid Company Ltd.), Kunming 650217, China*
[2]*Beijing Institute of Spacecraft System Engineering, Beijing 100094, China*
[3]*School of Computer Science, Guangdong University of Technology, Guangzhou 510006, China*

Correspondence should be addressed to Jiaying Li; 2112005161@mail2.gdut.edu.cn

Automated inspection using unmanned aerial vehicles (UAVs) is an essential means to ensure safe operations of the power grid. Defect detection for antivibration hammers on transmission lines in inspection imagery is one of the critical tasks for automated UAV inspection. It needs a machine interpretation system to automatically detect numerous inspection images. In this paper, a high-efficiency model based on Cascade RCNN (region-convolutional neural network) is proposed to detect antivibration hammer defects with reduced costs and speedier response, which applies in energy-efficient transmission line inspection systems. Firstly, to reduce computational costs, this study modifies the Cascade RCNN with a probabilistic interpretation to achieve the best trade-off between the inference time and average precision. Secondly, an antivibration hammer defect detector (AVHDD) model is proposed that uses a deep layer aggregation-based feature extraction network and a highly effective weighted bidirectional feature fusion network to replace the original ResNet and FPN on the modified Cascade RCNN to further enhance the model performance. Finally, a fine classification (FC) scheme for the types of antivibration hammer defects is proposed based on defect features to rationalize the model. The AVHDD reached an experimental mAP of 97.24% when IoU = 0.75, which is 2.93% higher than the original Cascade RCNN, and the defect recall was 98.9% while also significantly improving the inference speed. Moreover, the experimental results indicate that the overall performance of the proposed model is superior to typical models, confirming its suitability for energy-efficient transmission line inspection systems.

1. Introduction

The antivibration hammer is the key component of suspended transmission lines to suppress periodic vibrations and galloping in wires, as shown in Figure 1(a), where the hammer is enclosed in the rectangular boxes. Unfortunately, as the wire vibrates, it is repeatedly folded in the section near the suspension point for a long time, which further causes periodic fatigue broken strands, wire breakage, and tower collapse accidents. Thus, antivibration hammers play an irreplaceable role. However, they are prone to defective failures due to rusted metal and loose bolts, as shown in Figures 1(b) and 1(c), which prevent hammers from performing effectively as antivibration devices [1]. Therefore, it is vital to quickly identify and maintain defective antivibration hammers by performing detailed inspections.

Compared with unmanned aerial vehicles (UAVs), traditional manual inspection methods have problems such as long inspection cycles, low efficiency, great danger, poor ability to cope with complex terrain, and so on, which cannot easily meet the requirements of power grid operations and
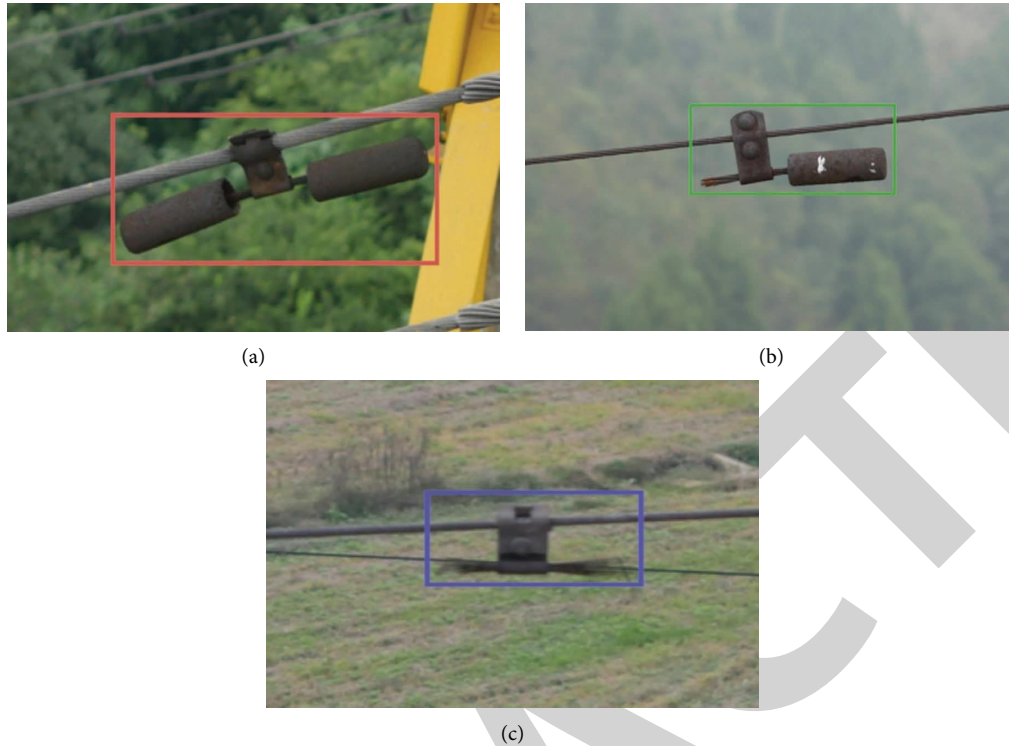
(a)

(b)

(c)

FIGURE 1: Antivibration hammer on transmission lines and their defective failures: (a) nondefective hammer, (b) partially defective hammer, and (c) completely defective hammer.

maintenance. As UAV inspections have the advantages of low costs, high efficiency, and a stronger ability to adapt to complex environments, various electric power research institutions have invested a great amount of human and material resources in recent years in the research of automated UAV inspection, including UAV control technologies with the goal of target identification and fault detection technologies in transmission line inspections from aerial images. The general trend is to replace manual operations with UAVs for intelligent inspections. Common UAVs capture visible light-based images or video using onboard cameras, which still require manual processing of the captured imagery to obtain transmission line status information. Manual judgment patrol images are prone to false detections or omissions and have high costs. To further improve the automation of UAV inspection and the detection performance of defective antivibration hammers, many scholars have proposed image-based transmission line antivibration hammer identification and defect-recognition methods to achieve rapid location and automated defect diagnosis. However, developing a high-efficiency antivibration hammer automation detection method for energy-efficient transmission line inspection systems is required to improve detection precision and reduce inference time.

2. Related Work

Current antivibration hammer automation detection methods are divided into three categories: traditional image recognition, machine learning, and deep learning.

Traditional image recognition algorithms perform antivibration hammer recognition and defect identification tasks by processing fixed features of the target (e.g., edges, colors, textures, and contours). Zhang et al. [2] used the Canny algorithm to extract edge information from images and performed circular and semicircular arc detection based on the center of mass, area, and contour of the segmented region to identify antivibration hammers. In reference [3], the method of vibration damper detection based on a random Hough transformation algorithm was proposed. In the work of reference [4], histogram equalization, morphological processing, and the red, green, and blue (RGB) color model were combined to process earthquake hammer images for corrosion defect detection. However, these algorithms rely on specific hammer angles and do not operate well in complex backgrounds. Traditional image recognition methods are susceptible to background interference, shooting distance, image brightness, and antivibration hammer angle. They rely on specific features, so they can only achieve good antivibration hammer detection in some situations. In brief, traditional image recognition methods for antivibration hammer recognition and defect identification have poor robustness and generalization abilities.

Some progress has been made in antivibration hammer recognition and defect identification based on machine learning. Miao et al.[5] extracted edges from images by using the wavelet modulus maximum algorithm. Selecting wavelet moments of edge images as the input features of the wavelet neural network achieved antivibration hammer identification through simulations. In reference [6], the Haar feature

was adopted to train the hierarchical AdaBoost classifier to recognize vibration dampers on transmission lines. Tian et al. [1] proposed an algorithm based on the fused chromatic aberration and radial basis function (RBF) neural network to identify faults in damper defects. In summary, machine learning-based methods often require complex algorithms to extract excellent features for use as inputs into intelligent classifiers. Therefore, this method is complex and requires a significant number of calculations.

Recently, machine vision technology based on deep learning has developed rapidly, and corresponding image object detection algorithms have achieved excellent performances. After the remarkable performance of AlexNet in the ImageNet competition in 2012, deep learning algorithms based on convolutional neural networks (CNN) have become the primary research direction for image classification and object detection due to the powerful ability of automatic feature extraction. In addition, large-scale public data sets and high-performance hardware processing systems have elevated object detection algorithms based on deep learning to new levels, which can be divided into one- and two-stage detectors. Specifically, one-stage detectors, as represented by the single-shot multibox detector (SSD) [7] and you only look once (YOLO) [8], have been used to detect antivibration hammers in transmission lines [9, 10]. In reference [10], the accuracies of antivibration hammer detection algorithms based on the SSD are significantly higher than machine learning-based methods used in reference [6]. However, while one-stage detectors have fast detection speeds, they have low detection precision for small targets.

In contrast, two-stage detectors are widely used to detect electrical components with higher detection accuracies, which include the region-convolutional neural network (RCNN), faster RCNN [11], and Cascade RCNN [12]. Among them, the antivibration hammer detection algorithm based on the faster RCNN has experienced a research upsurge [13]. Furthermore, reference [14] enhanced the feature extraction capabilities using a more powerful backbone as the feature extraction network of the faster RCNN and preprocessed the input image to reduce the negative impact of image quality inhomogeneity on the detection performance. Thus, the model not only can detect antivibration hammers but also can identify associated defects. However, the faster RCNN is prone to overfit if the intersection over union (IoU) threshold is set strictly at training, which has the problem of quality mismatch at inference [12]. To solve these issues, Bao et al. [15] proposed an improved Cascade RCNN model for antivibration hammer defect detection, which is better than other mainstream object detection algorithms in terms of detection precision. In summary, the application of Cascade RCNN for antivibration hammer defect detection has strong prospects. Nevertheless, in daily UAV inspection work, a large amount of image and video data is generated, necessitating a relatively high speed for transmission line inspection systems, which the slow detection speeds of standard two-stage detectors cannot easily meet.

To alleviate the slow detection velocity of two-stage detectors, we modified the Cascade RCNN with a probabilistic interpretation inspired by the literature [16] to realize rapid defect detection for antivibration hammers. Two advanced techniques are used to improve the model performance, including deep layer aggregation (DLA) [17] and bidirectional feature pyramid network (BiFPN) [18]. In addition, a fine classification (FC) scheme for defective antivibration hammers is proposed to rationalize the model following the intuitive concept that partially and completely defective hammers have significant feature differences. The above technologies are combined in an energy-efficient transmission line inspection system to detect antivibration hammer defects on transmission lines in a high-efficiency and low-energy manner.

The remainder of this paper is organized as follows. Section 3 details the proposed model and framework for antivibration hammer defect detection. Section 4 exhibits and discusses the experimental results. Finally, Section 5 presents the conclusion of this paper.

3. Methods

Antivibration hammer defect detection is a complex assignment that solves two key problems. First is antivibration hammer identification, which not only distinguishes between the foreground and background but also analyzes defects using the log-likelihood of the hammer. The second is antivibration hammer localization, which is handled using two-stage detectors through regression. The positive and negative samples are defined based on the IoU threshold. This threshold significantly impacts the training and inference, which is a hyperparameter that needs to be carefully selected. The IoU threshold of existing two-stage detectors is usually set to 0.5, which places a small constraint on the positive samples and gives the model prediction many approximations that result in generous noisy bounding boxes in the results. However, the excessive pursuit of high thresholds raises two problems. On the one hand, the number of positive samples decreases exponentially as the IoU threshold increases, which causes overfitting. On the other hand, using different thresholds in the training and inference phases leads to mismatches and degraded evaluation performances. When a low IoU threshold is chosen, more positive samples are produced, which benefits detector training and yet inevitably results in a large number of false detections during the inference process. The Cascade RCNN uses multiple stages for training, and the various stages have different IoU thresholds to define samples to train the model, which can solve this problem for two-stage detectors.

The structure of Cascade RCNN is shown in Figure 2, which is composed of a feature extraction module, a candidate region extraction module in the first stage, and a cascade detection module in the second stage. The ResNet-50 is used as the backbone network of the model to extract multiscale features from the input image. The extracted features are then subjected to feature fusion through the feature pyramid network (FPN) to obtain the final output feature maps with high-level semantic information. The candidate region extraction module is implemented primarily by the region proposal network (RPN) to generate coarse region proposals. In the cascade detection module,
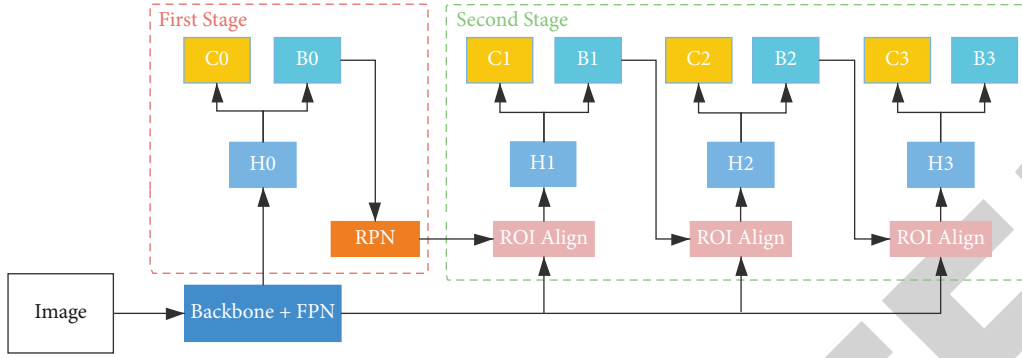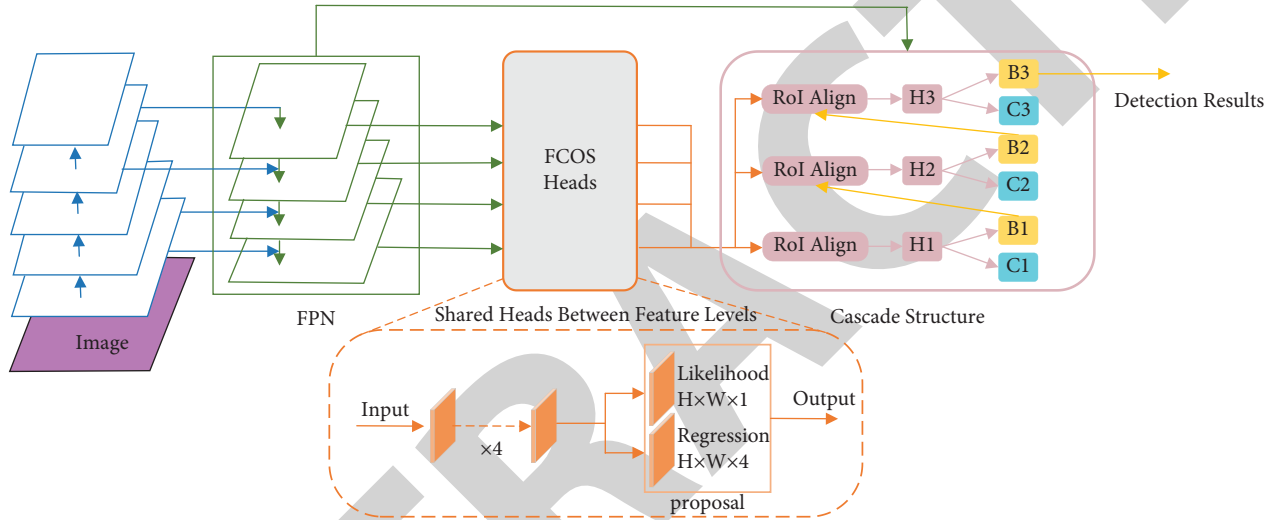
Figure 2: Structure of the Cascade RCNN.

Figure 3: Structure of the PI-Cascade RCNN.

these proposals are downsampled by the regions of interest (ROI) align layer and processed using a dedicated per-region head (H). Then, classification (C) and box regression (B) are performed, and the cascade structure is used to optimize the prediction results gradually.

3.1. Improved Cascade RCNN.

To date, most two-stage detectors use a relatively weak RPN for the first stage, which is designed to maximize the recall of a proposal and cannot produce an accurate likelihood. The abundance of proposals significantly reduces the speed of the model, and the recall-based RPN does not provide valid probabilistic interpretations as the one-stage detector [16]. Zhou et al. [16] used a one-stage detector to produce fewer however higher-quality proposals for the first stage of a two-stage detector. Specifically, the first stage infers the likelihood of the object rather than maximizing the recall. These likelihoods are then combined with the classification scores from the second stage to generate principle probability scores for the final detection. Their proposed probabilistic two-stage detector substantially reduces the number of proposals produced in the first stage (from 1 k to 256 k) and significantly reduces the inference time while improving the accuracy, which has a high practical application value.

To achieve rapid detection of antivibration hammer defects in transmission lines, a probabilistic interpretation based on the improved Cascade RCNN (PI-Cascade RCNN) was adopted here as motivated by the work of literature [16], as shown in Figure 3. It is noted that we did not change the feature extraction module and the cascade structure of the second stage of the original Cascade RCNN and used the fully convolutional one-stage object detector (FCOS) [19], which is an anchor-free one-stage detector with efficient shared heads instead of the RPN to generate high-quality region proposals in the first stage. Similar to the FCOS model, we apply shared detection heads to all feature maps generated from the FPN to reduce the model parameters and generate an object likelihood heatmap and bounding box regression map. To further speed up the process, we perform the likelihood branch while the regression branch shares the same convolution results from the four-layer structure. The likelihood branch in the detection head outputs a class-agnostic object likelihood $P(O = 1)$, which represents the probabilistic score of each feature point that belongs to the foreground on the feature map. The regression branch outputs the coarse relative distances l^*, t^*, r^*, and b^* of the location for the four edges of the bounding box. For each region proposal generated in the first stage, localization results are obtained by fine-tuning their four distance
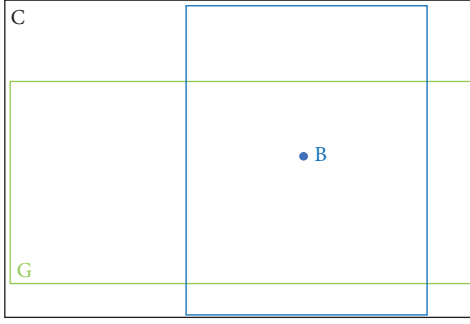
Figure 4: Diagram of the GIoU.

parameters in the second stage. Furthermore, the final category scores $P(C)$ are obtained by multiplying the probability scores $P(O = 1)$ in the first stage with the classification scores $P(C|O = 1)$ in the second stage as follows:

$$P(C) = P(C|O = 1)P(O = 1). \tag{1}$$

In the training phase of the PI-Cascade RCNN, we assign ground truth center annotations to specific feature levels using the same method and rules as the FCOS model. Then, the generalized intersection over union (GIoU) [20] is used to distinguish between the positive and negative samples instead of the IoU. A diagram of the GIoU is shown in Figure 4. For the bounding box B (blue rectangle) and ground truth G (green rectangle), the GIoU is evaluated as defined in

$$GIoU = \frac{|B \cap G|}{|B \bigcup G|} - \frac{\left|C - \left(B \bigcup G\right)\right|}{|C|}, \tag{2}$$

where B and G represent the area of the bounding box and the ground truth, respectively, and C is the area of the smallest enclosing convex region of B and G.

The loss function for the bounding box regression can be formulated as follows:

$$L_{reg} = 1 - GIoU. \tag{3}$$

During the training, a sample is defined as positive if its regression loss is less than 0.2. Here, we train the cascade classifier of the PI-Cascade RCNN using the maximum likelihood estimation. The classification loss for the annotated bounding box is described as follows:

$$L_{cls} = -\log\left[p^* P(C) + (1 - p^*)(1 - P(C))\right], \tag{4}$$

where p^* indicates whether the bounding box contains object 1 for yes and 0 otherwise.

The PI-Cascade RCNN requires a powerful first-stage detector to output accurate region recommendations. However, its performance is largely influenced by the quality of the feature maps from the feature extraction module. Therefore, we enhance the model performance by improving the feature extraction module and denote the final model as the antivibration hammer defect detector (AVHDD).

3.2. Feature Extraction Network of the AVHDD Model.

In deep learning-based object detection models, selecting a deeper backbone network can significantly improve the detection accuracy of the model. Some studies combine information between layers via skip connections to modestly increase network depth; the main structure of which is shown in Figure 5(a) [17]. However, such connections are still shallow, and the ability to learn feature information and retain shallow details is weak. The DLA is extended from shallow skip connections and achieves a close fusion of local features for each layer through concatenate operations, as shown in Figure 5(b). The aggregation structure of the DLA realizes the convergence of various stages from shallow to deep to achieve deeper semantic information sharing.

The DLA-34 [17] was used as the feature extraction network of the AVHDD to obtain deeper semantic information on antivibration hammer defects, which considers the speed and accuracy of the model. The specific structure of the DLA-34 is illustrated in Table 1.

As given in Table 1, the structure of the DLA-34 can be divided into six stages used to configure the basic blocks and convolution layers to be the same as ResNet-34. In addition, we use an aerial image of size $800 \times 800 \times 3$ as an input to the DLA-34 network, which outputs five feature maps of different scales as represented by $\{F2, F3, F4, F5, F6\}$. These were subsequently input to the FPN for multiscale feature fusion.

3.3. Feature Pyramid Network of the AVHDD Model.

For multiscale feature fusion, the FPN combines large-scale feature maps with the downsampled results of the small-scale feature maps in a top-down manner to increase the semantic information contained in the shallow feature maps to improve the detection capability of small objects. However, the FPN only performs one-way fusion, and the features of different layers are directly summed during fusion without considering their unequal contributions to the final output. Tan et al. [18] proposed the BiFPN to add skip connections to the traditional FPN while improving the fusion of top-down and bottom-up paths so the network can more fully fuse features of different scales without adding additional computational parameters. Nodes with only one input edge reduce the contributions of that node to the feature network. Here, we use the BiFPN as a feature pyramid network for AVHDD, as shown in Figure 6, which effectively fuses features at different scales and increases their information fusion at the same scale. After feature fusion, five feature maps are obtained, denoted as $\{P2, P3, P4, P5, P6\}$. Thus, high-quality regional proposals are generated by the FCOS heads.

3.4. Framework of Transmission Line Antivibration Hammer Defect Detection Method.

Figure 7 shows a framework diagram for the defect detection method of antivibration hammers for high-voltage transmission lines. The deep learning-based antivibration hammer image defect detection framework consists primarily of image processing, training the AVHDD model, and image detection. An antivibration
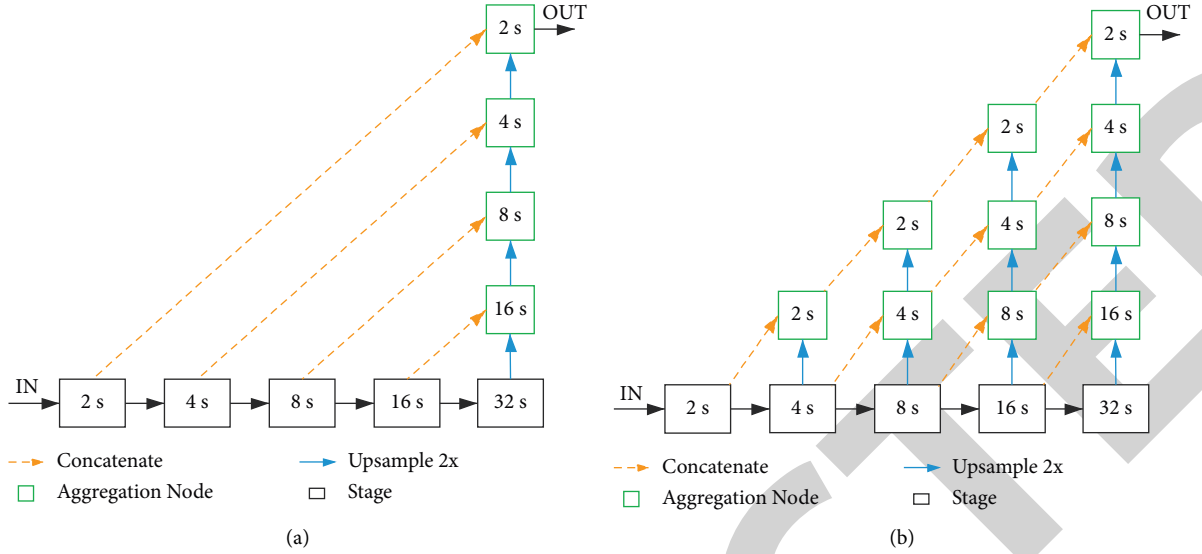
FIGURE 5: Comparison of the connection methods for the shallow and deep aggregation network structures: (a) shallow aggregation structure and (b) deep aggregation structure.

TABLE 1: Stage information of the DLA-34 structure.

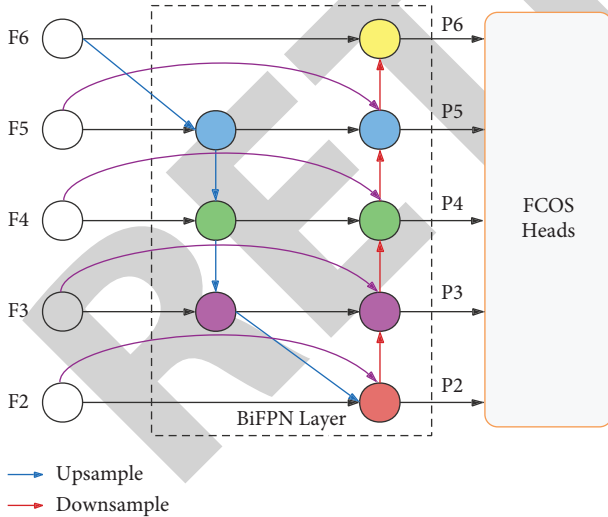| Stage	Input size	Stride	Output channel	Output name	Output size
1	$800 \times 800 \times 3$	2	16	—	$400 \times 400 \times 16$
2	$400 \times 400 \times 16$	4	32	$F2$	$200 \times 200 \times 32$
3	$200 \times 200 \times 32$	8	64	$F3$	$100 \times 100 \times 64$
4	$100 \times 100 \times 64$	16	128	$F4$	$50 \times 50 \times 128$
5	$50 \times 50 \times 128$	32	256	$F5$	$25 \times 25 \times 256$
6	$25 \times 25 \times 256$	64	512	$F6$	$13 \times 13 \times 512$

FIGURE 6: Structure diagram of the BiFPN.

hammer image data set is first established from the aerial images obtained from automated UAV inspection. During model training, we pretrain the proposed detection model using the COCO 2017 detection data set and later fine-tune it using the antivibration hammer data set to obtain the final AVHDD model. Once the antivibration hammer defect detection model is well trained, it can be used to detect

defects automatically and quickly in antivibration hammers on transmission lines as taken by UAVs. In Section 4.5, we define two classification schemes based on the defect features of the antivibration hammer (simple classification (SC) and FC schemes) and compare the AVHDD model's performance under each scheme.

4. Experiments and Results

First, the data set, evaluation indicators, compared models, and implementation details are all presented. The performance of our proposed AVHDD is compared to that of other models. The rationality of the scheme for FC of hammer defects is also verified.

4.1. Data Set. To the best of our knowledge, there are no publicly available data sets for antivibration hammer defect detection, so the images used in this experiment were taken from a UAV. However, due to the difficulty of capturing natural images that contain antivibration defective hammers, the number of antivibration hammer defect images in our data set is insufficient to train a reliable model. To obtain more antivibration hammer defect samples, we used Photoshop to simulate antivibration hammer defect samples by erasing some nondefective antivibration hammers under simple backgrounds. The schematic diagram to erase the nondefective antivibration
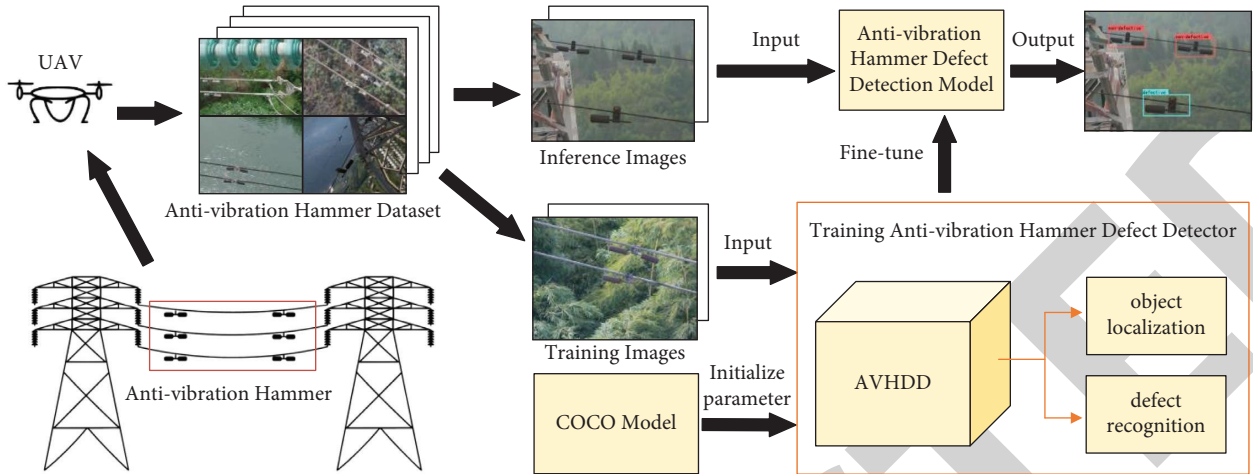
FIGURE 7: Framework diagram of the defect detection method of antivibration hammers for high-voltage transmission lines.
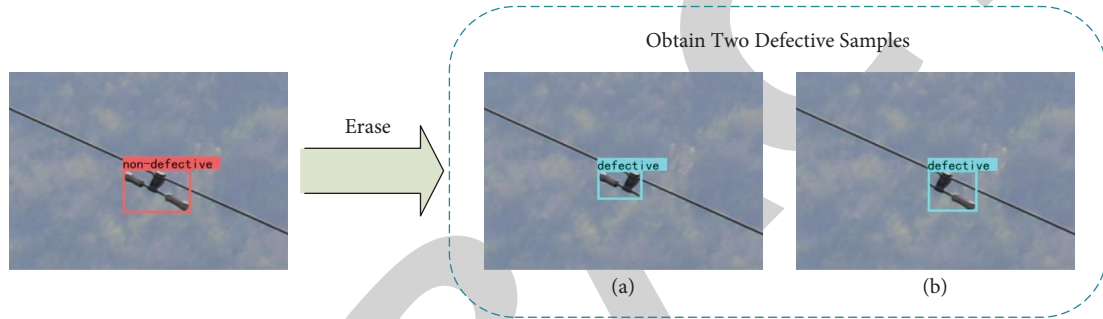

FIGURE 8: Schematic diagram of the process to erase parts of nondefective antivibration hammers.

hammers to generate the defective samples is shown in Figure 8.

To obtain robust models, our antivibration hammer data set is selected from some typical images and divided into four common conditions of antivibration hammer defect detection tasks, as shown in Figure 9. Specifically, Figure 9(a) gives two common types of antivibration hammers. Figure 9(b) exhibits different-sized antivibration hammers generated from various shooting distances. The antivibration hammers in the left image are much smaller than those in the right image. Two aerial images under different lighting conditions are shown in Figure 9(c). The left image shows that the antivibration hammer loses some of its features as the steel strands become transparent under strong light interference, while the right image was taken under normal lighting conditions. Figure 9(d) shows the samples of antivibration hammers under simple and complex backgrounds. Intuitively, antivibration hammers in complex backgrounds are more difficult to detect due to the significant interference from the background. Our antivibration hammer data set includes the four common conditions of antivibration hammer images in Figure 9 to enhance the sample diversity. Therefore, the models that perform well in this data set are robust and can be generalized for industrial applications of antivibration hammer defect detection to provide accurate hammer location and defect identification.

Most studies of antivibration hammer defect detection divided the sample categories into defective and nondefective based on the practical significance of whether the hammer has missing parts without considering differences in the features of partially defective and completely defective hammers. To rationalize the model, this paper proposes an FC scheme from the perspective of antivibration hammer features that classifies the antivibration hammer samples of the data set into three categories of nondefective, partially defective, and completely defective, as shown in Figure 1. The number of samples under different classification schemes is shown in Table 2.

4.2. Evaluation Indicators. To quantitatively measure the performance of the proposed model, we evaluate the detection speed using the inference time required to process an image, and the detection accuracy using the mean average precision (mAP), which is a comprehensive measure of detection accuracy under multiple recalls and is defined as:

$$mAP = \int_0^1 P(R)dR, \qquad (5)$$

where P represents the precision as the proportion of correctly predicted boxes to all predicted boxes, and R is the recall as the ratio of boxes that are correctly predicted in all ground truth. In (5), the mAP is equal to the area under the

(a)

(b)

(c)

(d)

FIGURE 9: Typical images of antivibration hammers divided into four common conditions: (a) two common types of antivibration hammers, (b) two different sizes of antivibration hammers, (c) two different lighting conditions of antivibration hammers, and (d) two different backgrounds of antivibration hammers.

Table 2: Number of samples under different classification schemes.

Classification scheme	Antivibration hammer samples	
	Category	Number
SC	Nondefective	1,880
	Defective	932
FC	Nondefective	1,880
	Partially defective	845
	Completely defective	87

precision-recall ($P − R$) curve and is valued between 0 and 1. In transmission line inspection processes, missed and false detections strongly influence inspection tasks. Thus, recall and false alarms (F) are also used as evaluation indicators to verify the model's reliability. The number of true positives (TPs), false positives (FPs), and false negatives (FNs) are used to calculate the precision, recall and false alarm in our experiments as follows:

$$P = \frac{\text{Num}(\text{TPs})}{\text{Num}(\text{TPs}) + \text{Num}(\text{FPs})},$$

$$R = \frac{\text{Num}(\text{TPs})}{\text{Num}(\text{TPs}) + \text{Num}(\text{FNs})}, \quad (6)$$

$$F = 1 − P.$$

A prediction box is defined as TP if the IoU between it and the ground truth is greater than the threshold and as FP if lower. An object missing detection is defined as FN.

4.3. Compared Models. The antivibration hammer defect detections are compared with some typical one- and two-stage detectors. One-stage detectors include anchor-based approaches (YOLOv4 [8] and RetinaNet [21]) and anchor-free approaches (FCOS [19] and CenterNet [22]). The faster RCNN [11] and Cascade RCNN [12] are the two-stage detectors used for comparisons in the experiment. To extract features from images, all compared models use ResNet-50 as the backbone network, except for YOLOv4, which uses CSPDarknet-53. In addition, all models are trained and tested using the antivibration hammer data set and the same experimental environment introduced in Section 4.4.

4.4. Implementation Details. All models are implemented based on the Detectron2 framework, which runs on an NVIDIA Tesla P100 GPU with 16 GB memory. Before training the proposed and compared models, we used the Albumentation toolkit to apply common stochastic data augmentation strategies, such as scaling the size and aspect ratio of the images, flipping them horizontally, and distorting their hue and saturation. The input images for all models need to be scaled to 800×800 for both training and inference. The stochastic gradient descent (SGD) optimization algorithm with a momentum of 0.9 and a weight decay of 0.0005 is used to adjust the parameters over the 150 training epochs. During model training, we set the learning

rate for the first 60 epochs to 10^{-4} and performed decay at epoch 60 and epoch 120 with a decay weight of 10^{-1}.

5. Results

For the best models obtained after training, we counted their mAP, average precision of each class, and time required to infer an image on the same data set individually. The performances of the models with the SC scheme are given in Table 3, where IoU = 0.5 and IoU = 0.75 represent the thresholds.

The quantitative results in Table 3 indicate that one-stage detectors have an absolute advantage in terms of running speed, while two-stage detectors perform relatively better in terms of detection accuracy. When the IoU threshold was 0.5, the YOLOv4 had the maximum mAP compared to the other models of 98.48%. However, when the IoU threshold was set strictly, the mAP of YOLOv4 showed a sharp decrease, as did the other one-stage detectors, while the mAP of the two-stage detectors showed a relatively stable performance under both thresholds. The RetinaNet has a modest performance, most likely because the majority of anchors generated in the experiment are redundant and cannot effectively improve the detection precision. The anchor-free based one-stage detectors FCOS and CenterNet have tremendous advantages in terms of inference speed, especially CenterNet, and yet they have weak accuracies and are not suitable for antivibration hammer defect detection tasks that require high accuracies. The ordinary faster RCNN performs worse than YOLOv4 in all aspects of the experiment. Adding the cascade structure in its second stage allows the mAP to reach 94.31% when IoU = 0.75, which is better than all other models. This is superior to all other models, however, at the cost of 10.5 ms more running time. Our proposed PI-Cascade RCNN model strikes the best trade-off between the inference time and average precision, which achieves a significant speedup while reducing the mAP by only 0.77% compared to the original Cascade RCNN. In brief, the detection accuracy of the Cascade RCNN outperforms all other compared models and has outstanding practical application value. After optimization via probabilistic interpretation, rapid antivibration hammer defect detection can be performed using only a slight reduction in the detection accuracy. In addition, when IoU = 0.5, the mAP of all models tends to saturate, making it difficult to compare the experimental results. Thus, using a more stringent threshold is beneficial for comparison experiments and provides more reliable results. In the following, we use an IoU threshold of 0.75 for the comparison experiments of the AVHDD model, and the associated experimental visualization results are shown in Figures 10 and 11.

Based on the observations in Figure 10, the detection precision of the AVHDD is significantly higher than that of PI-Cascade RCNN and Cascade RCNN, with an mAP of 95.74%. Furthermore, zooming in on the $P − R$ curve fragment for the recall interval of [0.7: 0.92] indicates that when the recall is greater than 0.92, the precision of the PI-Cascade RCNN and the unimproved Cascade RCNN begin to fall below 90% and then drop sharply. In contrast, the

TABLE 3: Performance of models on the same data set under the SC scheme.

| Model	IoU = 0.5			IoU = 0.75			Time (ms)
	Nondefective (%)	Defective (%)	mAP (%)	Nondefective (%)	Defective (%)	mAP (%)	
YOLOv4	98.64	98.32	98.48	86.96	92.65	89.81	48.9
RetinaNet	93.65	98.54	96.09	80.04	93.16	86.60	49.3
FCOS	94.06	96.15	95.10	81.75	90.38	86.06	44.9
CenterNet	93.39	96.61	95.00	80.96	86.18	83.57	40.7
Faster RCNN	93.33	97.96	95.64	81.92	94.97	88.32	62.0
Cascade RCNN	95.36	97.91	96.64	92.72	95.89	94.31	72.5
PI-Cascade RCNN	95.34	98.59	96.96	90.57	96.51	93.54	54.4
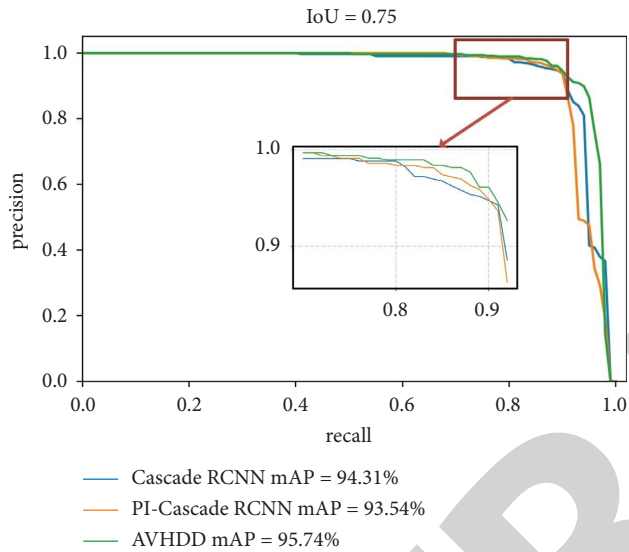
FIGURE 10: The $P - R$ curves for defect detection with three select models.
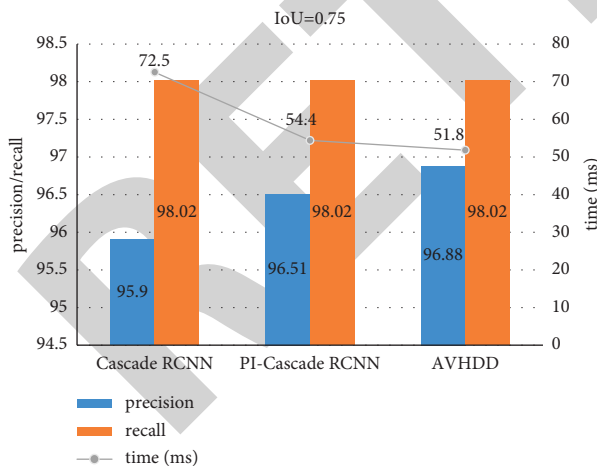
FIGURE 11: Effective analysis of the AVHDD model.

precision of the AVHDD is still above 90% and decreases more slowly. Specifically, the recall is much higher than the others. Therefore, we speculate that modifying the feature extraction module of the model using a combination of the DLA and BiFPN can effectively improve its robustness. To further exhibit the performance of the AVHDD model, we perform an efficiency analysis in terms of the precision and

TABLE 4: Recall, false alarm, and mAP of AVHDD under two classification schemes.

Classification Schemes	Category	Recall (%)	False alarm (%)	mAP (%)
SC	Nondefective	97.27	5.41	95.74
	Defective	98.02	3.12	
FC	Nondefective	97.73	3.54	97.24
	Partially defective	97.80	4.75	
	Completely defective	100.00	0.00	

recall of defective samples and the inference time, as shown in Figure 11.

Figure 11 shows that the AVHDD has a 0.98% higher defect detection precision than the Cascade RCNN when IoU = 0.75, and it takes only 51.8 ms to process an image of size 800×800. Combining the above experimental results, the inference speed of the AVHDD is comparable to YOLOv4 (48.9 ms), and with a much higher mAP (95.74% vs. 89.81%), which efficiently and accurately performs antivibration hammer defect detection.

Finally, we verify the effectiveness of the FC scheme based on the AVHDD model, and the experimental results are given in Table 4. The FC of defective samples into partially defective and completely defective improves the recall and false alarm rate of AVHDD, especially for completely defective samples, which are detected comprehensively and without errors under the FC scheme. Compared with the SC scheme, the model for defective antivibration hammers in the FC scheme reduces the average false alarms by 0.75%, with an average recall of 98.9%. Most importantly, using the FC scheme enhances the mAP to 97.24% without any increase in the inference time. In summary, it is reasonable to finely classify samples with antivibration hammer defects, which can effectively improve the detection accuracy of the model.

6. Conclusions

In this paper, an improved Cascade RCNN is applied for transmission line antivibration hammer defect detection based on UAV imagery, and an FC scheme of defective hammer samples is proposed to rationalize the model. First, we modify the Cascade RCNN with a probabilistic interpretation to produce fewer and higher-quality proposals

without affecting the precision. Then, the DLA backbone and BiFPN are then used to extract features to further improve the model performance, which was called the antivibration hammer defect detector (AVHDD).

Through experiments on our data set captured from a UAV, the proposed AVHDD model has a decisive advantage over other models in terms of detection accuracy, with a final mAP of 97.24% and an inference speed only slightly slower than YOLOv4. Moreover, finely classifying the antivibration hammer enhanced the recall of the model for defective samples to 98.9% without any increased time consumption. The fact that the energy consumption of systems normally depends on their execution time means that the proposed model is suitable and high-efficiency for energy-efficient transmission line inspection systems. In future work, we will simplify the model to achieve real-time detection without losing detection precision. At the same time, we will expand the defect detection scope of the model, such as detecting insulator defects, bird thorn defects, and foreign bodies on the transmission line.

Data Availability

All data, models, and code generated or used during the study are included within the submitted article.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Fangrong Zhou contributed to conceptualization; Gang Wen contributed to methodology and data; Guochao Qian and Yutang Ma contributed to software; Hao Pan and Jing Liu contributed to validation; Fangrong Zhou and Jiaying Li contributed to writing the original draft of the manuscript; Fangrong Zhou and Jiaying Li contributed to review and editing the manuscript; and Gang Wen contributed to supervision and funding acquisition. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

References

[1] Y. Tian, Y. Q. Wu, Y. Zhang, and X. B. Huang, "Fault identification of damper defect based on fused chromatic aberration and neural network," *Computer Technology and Development*, vol. 30, pp. 103–108, 2020.

[2] Fu S., Li W., and Zhang Y., "Structure-constrained obstacles recognition for power transmission line inspection robot, Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems," *IEEE*, pp. 3363–3368, 2006.

[3] X. J. Chen, Y. S. Wu, and L. Zhao, "Identification of OPGW vibration damper based on random Hough transformation," *Heilongjiang Electric Power*, vol. 32, pp. 1-2+5, 2010.

[4] W. Song, D. Zuo, B. F. Deng, H. B. Zhang, K. W. Xue, and H. Hu, "Corrosion defect detection of earthquake hammer for high voltage transmission line," *Chinese Journal of Scientific Instrument*, vol. 37, pp. 113–117, 2016.

[5] S. Y. Miao, W. Sun, and H. X. Zhang, "Intelligent visual method based on wavelet moments for obstacle recognition of high voltage transmission line deicer robot," *Robot*, vol. 32, no. 3, pp. 425–431, 2010.

[6] L. J. Jin, S. J. Yan, and Y. Liu, "Vibration damper recognition based on Haar-like features and cascade AdaBoost classifier," *Journal of System Simulation*, vol. 24, pp. 1806–1809, 2012.

[7] W. Liu, D. Anguelov, D. Erhan et al., "Ssd: Single shot multibox detector," in *Proceedings of the European conference on computer vision*, pp. 21–37, Amsterdam, The Netherlands, October 2016.

[8] 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.

[9] Z. A. Siddiqui and U. Park, "A drone based transmission line components inspection system with deep learning technique," *Energies*, vol. 13, p. 3348, 2020.

[10] S. Wang, *Research on Algorithms of Vibration Damper Detection on Power Line Image*, Beijing Jiaotong University, Beijing, China, 2017.

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, pp. 91–99, 2015.

[12] Z. Cai and N. Vasconcelos, "Cascade r-cnn: delving into high quality object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 6154–6162, Salt Lake City, UT, USA, June 2018.

[13] R. T. Jiao, H. X. Ni, and Z. Y. Wang, "Identification of transmission line damper based on Faster RCNN algorithm," *Journal of Changchun Institute of Technology(Natural Sciences Edition)*, vol. 22, pp. 38–43, 2021.

[14] H. Ni, M. Wang, and L. Zhao, "An improved Faster R-CNN for defect recognition of key components of transmission line," *Mathematical Biosciences and Engineering*, vol. 18, no. 4, pp. 4679–4695, 2021.

[15] W. Bao, R. Yangxun, L. Dong, Y. Xianjun, and X. Qiuju, "Defect detection algorithm of anti-vibration hammer based on improved cascade R-CNN," in *Proceedings of the 2020 International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI)*, pp. 294–297, EXCO Daegu, South Korea, November 2020.

[16] X. Zhou, V. Koltun, and P. Krähenbühl, "Probabilistic two-stage detection," 2021, https://arxiv.org/abs/2103.07461.

[17] F. Yu, D. Wang, E. Shelhamer, and T. Darrell, "Deep layer aggregation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2403–2412, Salt Lake City, Utah, June 2018.

[18] M. Tan, R. Pang, and Q. V. Le, "Efficientdet: scalable and efficient object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, Article ID 10790, Seattle, Washington, USA, June 2020.

[19] Z. Tian, C. Shen, H. Chen, and T. He, "Fcos: fully convolutional one-stage object detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9627–9636, Seoul, South Korea, October 2019.

[20] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: a metric and a loss for bounding box regression," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 658–666, Long Beach, California, June 2019.

[21] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, Venice, Italy, October 2017.

[22] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," 2019, https://arxiv.org/abs/1904.07850.