



ARTICLE

GFRF R-CNN: Object Detection Algorithm for Transmission Lines

Xunguang Yan^{1,2}, Wenrui Wang¹, Fanglin Lu¹, Hongyong Fan³, Bo Wu¹ and Jianfeng Yu^{1,*}

¹Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai, 201210, China

²University of Chinese Academy of Sciences, Beijing, 100049, China

³Jingwei Textile Machinery Co., Ltd., Beijing, 100176, China

*Corresponding Author: Jianfeng Yu. Email: yujf@sari.ac.cn

Received: 27 August 2024 Accepted: 30 October 2024 Published: 03 January 2025

ABSTRACT

To maintain the reliability of power systems, routine inspections using drones equipped with advanced object detection algorithms are essential for preempting power-related issues. The increasing resolution of drone-captured images has posed a challenge for traditional target detection methods, especially in identifying small objects in high-resolution images. This study presents an enhanced object detection algorithm based on the Faster Region-based Convolutional Neural Network (Faster R-CNN) framework, specifically tailored for detecting small-scale electrical components like insulators, shock hammers, and screws in transmission line. The algorithm features an improved backbone network for Faster R-CNN, which significantly boosts the feature extraction network's ability to detect fine details. The Region Proposal Network is optimized using a method of guided feature refinement (GFR), which achieves a balance between accuracy and speed. The incorporation of Generalized Intersection over Union (GIOU) and Region of Interest (ROI) Align further refines the model's accuracy. Experimental results demonstrate a notable improvement in mean Average Precision, reaching 89.3%, an 11.1% increase compared to the standard Faster R-CNN. This highlights the effectiveness of the proposed algorithm in identifying electrical components in high-resolution aerial images.

KEYWORDS

Faster R-CNN; transmission line; object detection; GIOU; GFR

1 Introduction

The global expansion of power transmission networks has led to an increase in electrical incidents, causing substantial economic losses and safety risks. Traditionally, inspections of these infrastructures were manual, costly, and inefficient. However, the integration of unmanned aerial vehicles (UAVs) with advanced deep learning detection techniques has revolutionized this process [1]. UAVs can navigate complex environments and capture high-resolution images of electrical components, essential for maintenance and safety management. Further enhancing this innovation, the advent of 5G and 6G technologies has significantly improved UAV inspection capabilities, particularly for power transmission lines. These technologies have boosted the clarity and resolution of drone-captured images, greatly enhancing the potential for accurately identifying small but crucial components, such



as screws. This advancement ensures even the minutest details are captured, leading to more precise and reliable inspections and marking a significant step in enhancing the safety and efficiency of power transmission infrastructure.

The high-definition images captured by drones often contain small yet crucial objects like screws, which, due to their limited size in the image, risk being overlooked by traditional detection algorithms. Additionally, transmission lines are often situated amidst diverse and challenging environments such as mountains, rivers, grasslands, and under varying lighting conditions. These factors complicate the accurate detection of line components, presenting a significant challenge in ensuring effective monitoring.

In this research, the Faster R-CNN [2] framework is enhanced to address the challenges of detecting small objects in high-resolution images. The approach commences with the implementation of a copy-paste [3] data augmentation technique, creating a comprehensive dataset that includes insulators, shock hammers, and screws. Subsequently, the standard VGG16 [4] network in the Faster R-CNN's feature extraction module is replaced with an advanced ResNet50 network, integrated with a Feature Pyramid Networks (FPN) [5] multi-scale pyramid structure. This modification significantly improves the backbone network's ability to extract detailed information. Furthermore, a guided feature refinement method is employed to expedite the inference speed of the feature maps. The bounding box regression mechanism is also refined using GIOU [6] and ROI Align [7]. Collectively, these enhancements ensure that the improved Faster R-CNN not only maintains high detection precision but also increases the speed of the detection process, rendering it highly effective for target detection in high-resolution imagery.

This research introduces several key advancements in the field of power transmission line component detection using an enhanced Faster R-CNN framework:

1. **Development of a High-Precision Detection Algorithm:** A detection algorithm specifically for critical components of power transmission lines is refined. This is achieved by employing a custom high-resolution image dataset for data augmentation, leading to a model that not only demonstrates high precision but also maintains a relatively swift detection rate. This makes it particularly effective for identifying transmission line components in high-resolution images.

2. **Backbone Network Enhancement:** In the improved Faster R-CNN, the traditional VGG16 network is replaced with an advanced ResNet50 network, integrated with an FPN multi-scale feature pyramid structure. This adaptation significantly enhances the network's ability to detect a diverse range of target sizes, including small, medium, and large objects, thereby elevating its overall capacity for feature extraction.

3. **Region Proposal Network (RPN) Optimization for Small Target Detection:** To improve the detection accuracy of small targets such as screws, the P2 layer is retained in the Region Proposal Network (RPN), despite its increased computational demands. A guided feature refinement method is employed that transfers partial positional information of small screws from the P3 layer, aiding in inference in subsequent layers. This strategy not only maintains prediction accuracy but also boosts inference speed at the P2 layer.

4. **Bounding Box Regression Refinement:** The original Intersection Over Union (IOU) structure is replaced by Generalized IOU (GIOU). GIOU considers the non-overlapping areas between candidate priors and actual objects, addressing the limitations of the traditional IOU. Furthermore, ROI Align is utilized instead of ROI Pooling to correct dual quantization errors, leading to more accurate bounding box regression.

2 Related Work

2.1 Development of Transmission Line Target Detection Algorithms

The traditional object detection methods, including the Viola-Jones (VJ) framework [8], Histogram of Oriented Gradients (HOG), and Deformable Part Models (DPM) [9], concentrate on morphological characteristics and appearance patterns for initial image analysis. These methods involve manual extraction of features for identifying and locating objects [10]. While straightforward in computation, these traditional techniques often suffer from slow processing speeds. Additionally, they are vulnerable to disruptions caused by complex backgrounds and fluctuating lighting conditions, which can significantly hinder their effectiveness in dynamic or challenging settings.

The landscape of object detection algorithms has been revolutionized by the advent of deep learning, leading to the gradual obsolescence of traditional methods. Currently, the most prominent object detection models are categorized into two primary approaches: region proposal-based two-stage models and regression-based one-stage models. Region proposal-based models, such as Region Convolutional Neural Networks (R-CNN) [11], Fast R-CNN [12] and Faster R-CNN, have been pivotal in this evolution. Faster R-CNN, in particular, introduced the RPN and the concept of Anchors. These models use bounding boxes of varying aspect ratios to precisely locate objects, significantly improving training efficiency. On the other hand, one-stage models like YOLOv2 [13], YOLOv3 [14], YOLOv4 [15] and Single Shot MultiBox Detector (SSD) [16] approach detection as a regression problem. They directly predict object targets and their bounding box attributes from image pixels. While these one-stage models may sacrifice some accuracy for speed, they offer a substantial increase in processing velocity. YOLO, especially, is known for its rapid detection capabilities. The SSD model combines YOLO's regression approach with the anchor mechanism from Faster R-CNN, creating a hybrid that leverages the strengths of both methodologies.

In the evolving field of power transmission system inspections, the integration of advanced object detection algorithms has markedly enhanced the precision and efficiency of identifying critical components in high-resolution images captured by UAVs. This progress is crucial for improving inspection accuracy. Xu et al. [17] proposed an efficient YOLOv3-based method for detecting transmission line defects, optimizing the model to achieve real-time detection in UAV inspection videos while maintaining high accuracy, and the enhancement of the SSD algorithm using MnasNet by Liu et al. [18] to increase detection speed for insulators and spacers. Bao et al. [19] modified YOLOv4 by incorporating a Parallel Mixed Attention module, focusing the network's attention on critical areas in complex backgrounds. Zhao et al. [20] employed a Feature Pyramid Network with Faster R-CNN for precise insulator localization in intricate backgrounds, followed by segmentation. Further, Zhang et al. [21] enhanced the Faster R-CNN model for detecting bird nests on transmission lines by incorporating Inception-v3 and a dual attention mechanism, achieving substantial improvements in recognition accuracy and detection speed. Chen et al. [22] improved the YOLOv5s model by integrating the CBAM and focal loss function, significantly enhancing detection accuracy and performance for key components in complex transmission line environments. Liu et al. [23] combined YOLOv3 with a Cross Stage Partial Network to detect insulator faults against complex backgrounds, achieving performance comparable to YOLOv4 with improved mAP. Liu et al. [24] addressed class imbalance in transmission line images through data augmentation, optimizing the YOLOv4 algorithm to achieve over 90% accuracy and recall. Zhai et al. [25] proposed a hybrid knowledge region-based convolutional neural network, effectively extracting spatial relationships of transmission line fittings and improving detection performance for multiple joints on power lines. While existing models like SSD, YOLO, and Faster R-CNN have laid a strong foundation in object detection, their application in power line

inspections faces limitations, particularly in accurately detecting small, critical components under complex environmental conditions. The GFRF R-CNN model introduces an innovative GFR method, enhancing detection capabilities beyond the constraints of these traditional models. By integrating a more sophisticated feature extraction mechanism tailored for high-resolution imagery, GFRF R-CNN achieves superior precision in identifying small-scale electrical components, a crucial advancement for maintaining the integrity and safety of power systems. This novel approach not only improves upon the detection accuracy but also optimizes processing speed, addressing the critical need for efficient real-time analysis in UAV-based inspection systems.

2.2 The Algorithm Flow of Faster R-CNN

In the Faster R-CNN algorithm, the initial stage involves image preprocessing and feature extraction, where the image is processed through a network like VGG16 to create a feature map. This map is then used in the Region Proposal Network, where anchors representing potential object locations are generated for each cell. A 3×3 sliding window processes these anchors, producing vectors that feed into both a box-regression layer and a box-classification layer. The regression layer predicts bounding box adjustments, while the classification layer assesses whether anchors are foreground or background. The RPN then generates candidate boxes based on these predictions, which are mapped onto the feature map to obtain corresponding feature matrices. Finally, each matrix undergoes ROI pooling to standardize its size, and after flattening, it passes through fully connected layers. A softmax classifier then produces the final prediction results, completing the process. The structure of Faster R-CNN is shown in Fig. 1.

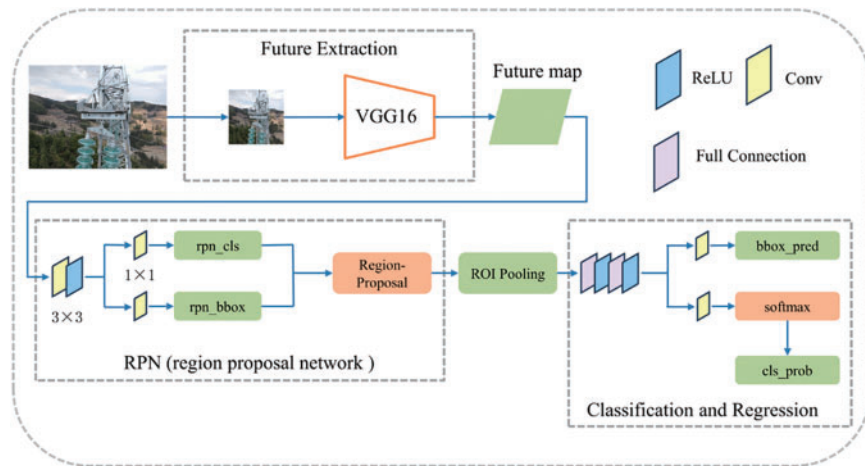


Figure 1: The structure of Faster R-CNN

2.3 The Loss Function of the Faster R-CNN

Following the segregation of positive and negative samples for training, the loss within the RPN segment is computed. This loss encompasses both classification loss and bounding box regression loss. The formula for the RPN multi-task loss is presented as Eq. (1):

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (1)$$

Here, L_{cls} represents the classification loss and L_{reg} denotes the regression loss. The variable p_i is the predicted probability of an anchor being an object, and p_i^* is the ground truth label (1 for positive, 0 for negative). The term t_i is the vector of predicted bounding box parameters, while t_i^* is the vector of ground truth bounding box parameters. N_{cls} and N_{reg} are normalization terms, and λ is a balancing parameter that adjusts the relative weight of the classification and regression losses. The classification loss, L_{cls} is computed using Binary Cross Entropy (BCE), which is defined as Eq. (2):

$$L_{cls} = -[p_i^* \log(p_i) + (1 - p_i^*) \log(1 - p_i)] \quad (2)$$

The regression loss, L_{reg} utilizes the Smooth L1 loss, which is less sensitive to outliers than the L2 loss and is defined as Eqs. (3) and (4):

$$L_{reg}(t_i, t_i^*) = \sum_i smooth_{L_1}(t_i - t_i^*) \quad (3)$$

$$Smooth(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & |x| \geq 1 \end{cases} \quad (4)$$

In the Faster R-CNN framework, the loss function for the second part, following the RPN, also consists of two components: a classification loss and a bounding box regression loss. This part of the model refines the proposals generated by the RPN and performs the final object classification and bounding box regression. The loss function can be expressed as Eq. (5):

$$L = L_{cls}(P, u) + \lambda L_{reg}(t^u, v) \quad (5)$$

where $L_{cls}(P, u)$ is the classification loss, computed as the cross-entropy loss over the softmax probabilities P for the correct class label u . The regression loss $L_{reg}(t^u, v)$ is calculated using the Smooth L1 loss, similar to the RPN stage, where t^u represents the predicted bounding box adjustments for class u , and v denotes the ground truth bounding box parameters. The term λ serves as a balancing weight between the classification and regression losses.

In the Faster R-CNN algorithm, the tasks of classification and regression are intricately linked. The classification loss function plays a crucial role in accurately identifying the object class, while the regression loss function focuses on precisely refining the bounding box's spatial positioning relative to the object. During the training phase, these two aspects are simultaneously optimized through a combined loss function. This optimization process is key to improving the model's overall performance, ensuring not only effective object detection but also accurate localization.

3 Proposed Method

3.1 Improvement of the Faster R-CNN

In this research, significant enhancements have been made to the Faster R-CNN framework to improve its performance in high-resolution image analysis. The primary improvement involves replacing the original VGG16 feature extraction module with an advanced ResNet50 model, which is seamlessly integrated with an FPN. This combination forms the backbone of the revamped network, significantly enhancing its feature extraction efficiency. With the increased dimensions of input images, the RPN has been tailored to process feature maps from layers P3 to P6, ensuring efficient inference times for larger image sizes.

However, a challenge was encountered in precisely detecting smaller objects, such as screws, within this framework. To overcome this, the P2 layer has been strategically incorporated into the network architecture. Furthermore, a guided feature refinement method has been implemented at the P2 layer, specifically designed to accelerate the inference process. This approach markedly improves the network's ability to detect and process smaller objects with high accuracy, a critical requirement for analyzing high-resolution images.

The updated structure of the Faster R-CNN network is detailed in a figure included in the documentation. This illustration highlights the extensive enhancements integrated into the existing framework, showcasing the commitment to advancing object detection technology in complex and high-resolution imaging scenarios. The structure of the enhanced Faster R-CNN, referred to as GFRF R-CNN, is depicted in Fig. 2.

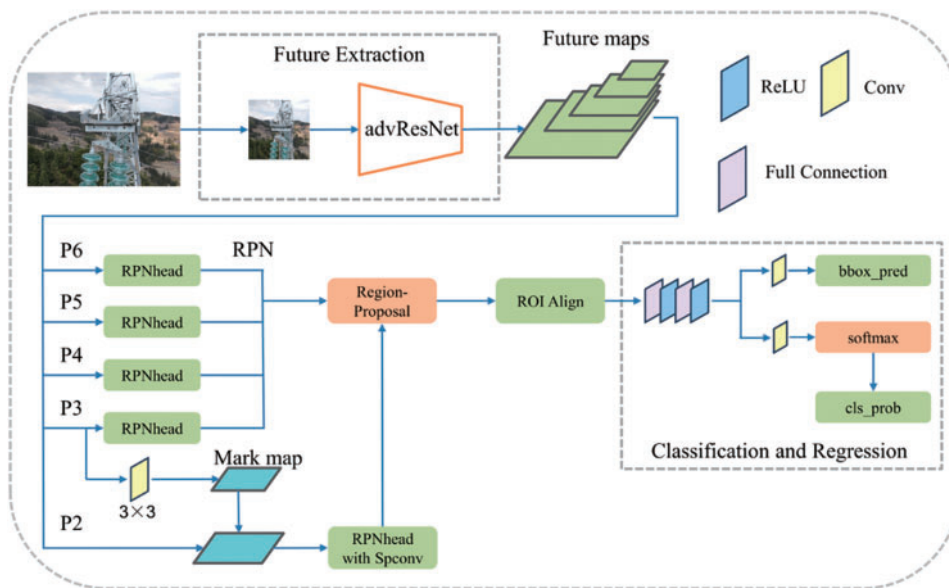


Figure 2: The structure of GFRF R-CNN

3.2 Improved Backbone Network

This research introduces the AdvResNet50, a refined feature extraction network within the Faster R-CNN framework, specifically tailored to address the challenges of complex backgrounds and variable lighting conditions in power line imagery. The modifications to the original residual network (ResNet) [26], now termed AdvResNet, are designed to enhance the detection of high-resolution components in such challenging environments. The modifications made are illustrated in Fig. 3. Key improvements include:

The residual network's bottleneck structure has been modified. It starts with a 1×1 convolution that expands the feature map's dimensionality, followed by a 3×3 depth-wise convolution [27] for feature extraction. The process concludes with another 1×1 convolution, returning to the feature map's original dimensions. This design allows the 3×3 convolution to handle a larger number of channels, facilitating the extraction of a more diverse range of information. The depth-wise convolution ensures that the parameter count remains manageable.

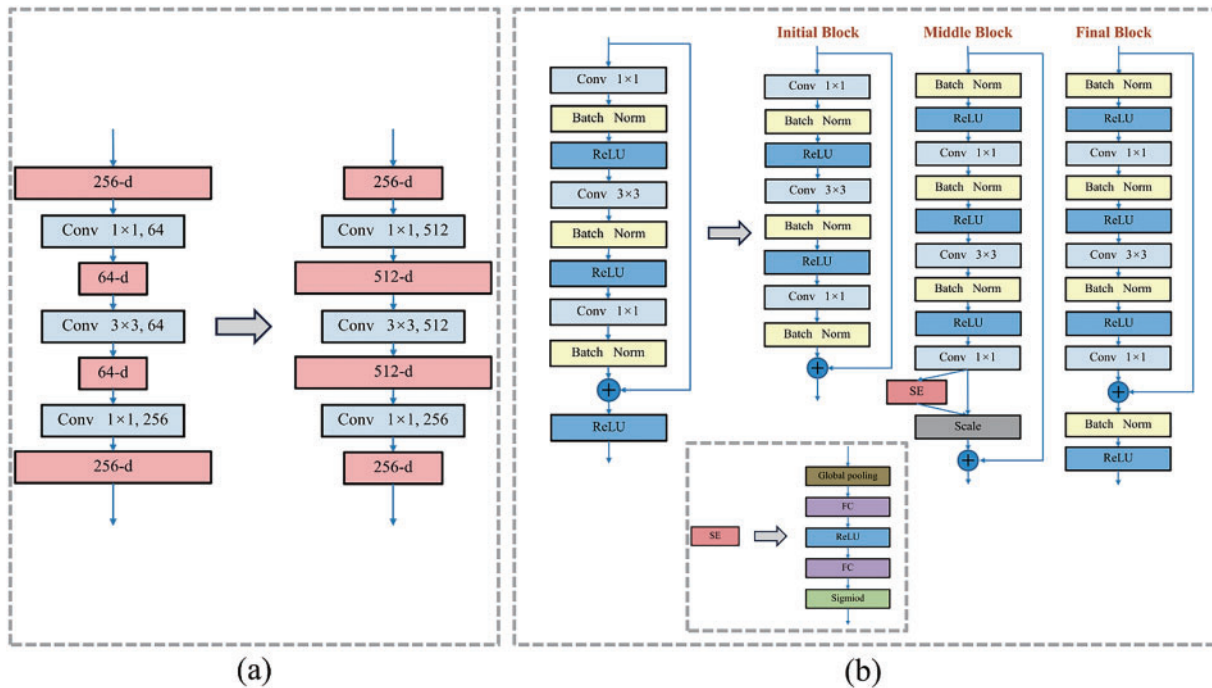


Figure 3: AdvResNet: (a) Modification of bottleneck (b) Initial, middle and final block

The architecture diverges from the traditional method of stacking identical modules by incorporating three distinct types: initial, middle, and final. Each stage of the residual network is composed of an initial block, several middle blocks, and a final block. The initial block omits the ReLU activation from the main path, allowing negative signals to propagate. The middle block repositions the ReLU layer to the beginning of the main branch, altering the sequence of components without increasing the network’s parameter complexity. Additionally, in the middle block, an SE (Squeeze-and-Excitation) [28] module has been added, employing channel attention mechanisms to enhance the model’s sensitivity to channel features, which is beneficial for extracting key information. The final block, building on the middle block, adds a Batch Normalization (BN) and ReLU layer to the main path. This ensures stage-wise information normalization and enhances the network’s capacity for non-linear learning.

These strategic enhancements, in conjunction with the FPN, form a powerful feature extraction system. This system is adept at detecting high-resolution components in power line images, effectively navigating the complexities of diverse backgrounds and lighting variations.

3.3 GFR Method

In this research, a novel GFR method is introduced to enhance the detection of small objects in high-resolution images, particularly focusing on the P2 layer of the RPN within the Faster R-CNN framework. This method is specifically designed to address the challenges posed by small objects, such as screws, which are often overlooked due to their minimal presence in high-resolution images.

The GFR method operates within the context of the FPN, where feature maps of varying resolutions are generated. For an input image with dimensions $H \times W$, the size of the FPN feature

maps is represented as $P = \left\{ P_k \in \mathbb{R}^{\frac{H}{2^k} \times \frac{W}{2^k} \times C} \right\}$, with k indicating the pyramid level. In this setup, small objects are typically more discernible in the high-resolution, lower-level feature maps, such as P2.

To enhance comprehension of the GFR mechanism, an analogy is introduced that likens the process to human visual perception at varying distances. High-level feature maps within the proposed model are compared to the expansive view observed from a distance, facilitating the identification of general areas where small objects, such as distant landmarks, might be situated. This panoramic perspective allows for the approximate locations of small targets to be determined within a broad landscape.

Conversely, low-level feature maps are analogous to the detailed view attained when closer proximity to an object is achieved. Such proximity affords a thorough and detailed examination, facilitating the precise identification of specific features of small objects. The GFR mechanism employs this principle by initially harnessing the ‘distant’ view offered by high-level feature maps to identify areas of interest broadly. Subsequently, a ‘zoom in’ approach is adopted with low-level feature maps for a detailed and refined detection process, accurately pinpointing the exact positions of small-scale objects within these preliminarily identified regions.

The GFR method starts by predicting approximate locations of small objects on the P3 feature map. This step is crucial for highlighting areas of interest where small objects, such as screws, are likely to be found. By determining these potential locations, the GFR method efficiently allocates computational resources, focusing on areas with a high likelihood of containing small objects. This targeted approach reduces the overall processing load, enabling more accurate and efficient small object detection.

The GFR method employs a sophisticated mapping technique to identify cells containing small objects on the P3 layer, which are then marked distinctly on a ground truth map. This process involves a binary classification challenge framed as a pixel-level binary cross-entropy loss, which effectively discriminates between cells containing small objects and those that do not. The refined map T generated for the P3 layer plays a crucial role during inference, identifying cells marked as containing small objects and mapping these onto a corresponding sparse map on the P2 layer. This map acts as a selective filter, preserving marked areas and discarding others, leading to a sparse feature map that undergoes further processing through an RPN head equipped with sparse convolution.

In the context of the Faster R-CNN framework, the GFR method involves a specialized module, functioning parallel to the RPN head, which is instrumental in predicting the approximate locations of small objects.

In our research, the GFR method is adeptly applied to the P3 feature layer, utilizing dual convolution layers to create a sophisticated feature map, $T \in \mathbb{R}^{\frac{H}{8} \times \frac{W}{8}}$. This map, $T(i, j)$, effectively measures the probability of small, otherwise undetectable objects at the P3 layer, existing within each grid cell (i, j) . These objects are defined as those smaller than the P3 layer’s smallest anchor. For every small object detected, its central coordinates (x, y) are mapped to the P3 layer’s corresponding grid cell. This cell is then distinctly marked as 1 on the ground truth map, T^* , with all other cells marked as 0. The schematic representation of the GFR method is depicted in [Fig. 4](#).

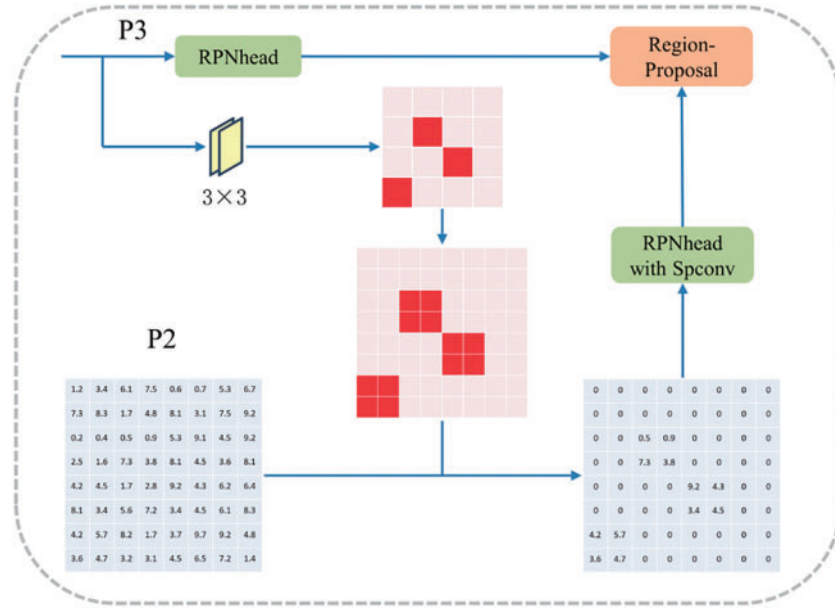


Figure 4: GFR method in RPN

A novel loss function, L_{total} , is introduced for training. It ingeniously combines the original Faster R-CNN RPN loss (L_{FR}) with the GFR module loss (L_{GFR}), moderated by a coefficient λ , which is described as Eq. (6):

$$L_{total} = L_{FR} + \lambda L_{GFR} \quad (6)$$

The L_{GFR} is ingeniously crafted as a binary cross-entropy loss, framing the output as a pixel-level binary classification challenge. The formula for L_{GFR} is presented as Eq. (7):

$$L_{GFR} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(y_i^*) + (1 - y_i) \log(1 - y_i^*)] \quad (7)$$

In this equation, N denotes the total grid cell count, y_i is the ground truth (0 or 1), and y_i^* the predicted probability.

In the GFR method, the refined map T , generated for the P3 layer, plays a crucial role during inference. This map identifies cells containing small objects, marked as 1. These identified cells are then mapped onto a corresponding sparse map, $T' \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4}}$, on the P2 layer. T' acts as a selective filter on the P2 layer, preserving areas marked 1 and nullifying others. This selective process results in a sparse feature map on the P2 layer, which is then processed through an RPN head equipped with sparse convolution [29].

This methodological innovation not only streamlines the detection process but also markedly enhances the precision of identifying small objects in intricate high-resolution images. The incorporation of the GFR method into the Faster R-CNN framework represents a substantial advancement in object detection technology, especially in the realm of high-resolution imaging challenges.

3.4 Align Pooling

In this study, the feature extraction process within the Faster R-CNN framework has been refined, with a focus on improving the handling of complex backgrounds and variable lighting conditions in power line imagery. A key enhancement addresses the pixel misalignment issue commonly encountered in traditional ROI Pooling. This problem typically arises from converting image coordinates to feature map coordinates and then to ROI feature coordinates, often leading to quantization errors due to the necessity of rounding floating-point numbers to integer pixel coordinates.

To address this challenge, ROI Align, as introduced in the Mask R-CNN paper, has been incorporated. This technique preserves the original floating-point pixel coordinates, thereby avoiding the dual rounding errors inherent in traditional ROI Pooling. ROI Align utilizes bilinear interpolation, an advanced image scaling method that calculates the pixel value of a virtual point based on the values of the four nearest actual points surrounding it. The formula for bilinear interpolation is detailed in Eq. (8):

$$f(x, y) = f(Q_{11})(x_2 - x)(y_2 - y) + f(Q_{21})(x - x_1)(y_2 - y) \\ + f(Q_{12})(x_2 - x)(y - y_1) + f(Q_{22})(x - x_1)(y - y_1) \quad (8)$$

In this formula, Q_{11} , Q_{21} , Q_{12} , Q_{22} represent the coordinates of the four adjacent points, x, y are the coordinates of the virtual point, and $f(x, y)$ is the calculated pixel value of the virtual point. This method ensures a more precise and aligned feature extraction from the region of interest, significantly boosting the accuracy of object detection models in complex imaging scenarios.

3.5 GIOU

In this section, a critical limitation in object detection metrics has been addressed by replacing the traditional IOU metric with the GIOU in the Faster R-CNN algorithm. The IOU metric, while widely used, has notable shortcomings. It measures the overlap ratio between the predicted bounding box and the actual object but fails to consider the size and shape of objects and bounding boxes. Moreover, IOU struggles with two specific issues: (1) it yields a zero value when there's no overlap between the predicted bounding box and the actual object, making it non-optimizable, and (2) it cannot differentiate between various alignments of bounding boxes with respect to the ground truth, even when they have the same loss values but different IOU scores.

GIOU addresses these limitations by extending the concept of IOU to include the non-overlapping areas between the bounding box and the actual object. This enhancement allows for a more nuanced and accurate assessment of the bounding box's precision, especially in scenarios where the traditional IOU would yield a zero value or fail to differentiate between various alignments.

By integrating GIOU into the Faster R-CNN framework, the model has been enabled to more effectively optimize bounding boxes during training. This leads to an improved performance of the Faster R-CNN model in object detection tasks. The adoption of GIOU represents a significant advancement in object detection technology, providing a more comprehensive and accurate evaluation of bounding box predictions.

4 Experiments

The experimental evaluation of GFRF R-CNN model was carried out on a robust computing system operated on Windows 11 and utilized Python 3.8 as the programming language. The hardware configuration included a Xeon Gold 5218R CPU and 32 GB of RAM, supported by an RTX2080Ti

GPU with 16 GB of video memory, ensuring efficient processing and computation capabilities. The software environment was anchored by CUDA (Compute Unified Device Architecture) version 11.8, a parallel computing platform and application programming interface model created by NVIDIA, based in Santa Clara City, CA, USA. Additionally, the model utilized PyTorch version 1.11.2, an open-source machine learning library developed by Facebook, located in Park, CA, USA.

To rigorously test and train the model, a dataset was compiled using a drone equipped with high-resolution imaging capabilities. This dataset was crucial in evaluating the performance and accuracy of the GFRF R-CNN. By utilizing high-resolution aerial imagery, the model's applicability and effectiveness were ensured to be tested in real-world scenarios, simulating practical applications where such advanced object detection capabilities are essential.

4.1 Dataset Preprocessing

To effectively simulate diverse lighting conditions that might occur at different times of the day or under varying weather scenarios, a technique known as color jittering was employed in the study. This method involves adjusting key image attributes such as brightness, saturation, and contrast. The mathematical basis for this transformation is encapsulated in Eq. (9):

$$g(x) = a \cdot f(x) + b \quad (9)$$

In this equation, $f(x)$ represents the original pixel values, while $g(x)$ denotes the adjusted pixel values. The factor a serves as the gain factor, influencing the image's contrast, and b acts as the bias factor, adjusting the image's brightness.

The processed dataset was annotated using LabelImg, a graphical image annotation tool. Additionally, the data augmentation process was refined by modifying the copy-paste method. Initially, two images were randomly selected from the original dataset. These images were subjected to standard scale jittering and random horizontal flipping. The scaling factors were confined within the range of 0.8 to 1.25, a notable shift from the original method that used large-scale jittering with scaling factors between 0.1 and 2.0. This change was implemented after observing that large-scale jittering adversely affected the accuracy of detecting small objects. By adopting standard scale jittering, an overall improvement in detection accuracy was achieved.

The final step in this refined data augmentation process involved copying bounding boxes containing specific objects, such as insulators and vibration dampers, from one image and pasting them onto another. This technique further enhanced the robustness of our model by providing a more diverse range of training data, crucial for improving the model's performance in real-world scenarios. The process for dataset generation is depicted in Fig. 5.

In the dataset acquired for this study, a subset of images captured during foggy and rainy conditions was included to ensure a representation of diverse environmental scenarios. Following data augmentation processes, a total of 3769 images were generated. These images were then systematically divided into training, validation, and test sets using a 6:3:1 ratio, ensuring a comprehensive evaluation framework for assessing the model's robustness across varying weather conditions.

In addition to the color jittering and copy-paste methods previously mentioned, our preprocessing regimen included the application of fog and motion blur effects to a subset of our images. These effects were implemented to simulate less-than-ideal weather conditions and potential blurring due to UAV movement, respectively. Fog addition mimics the atmospheric conditions that may affect the clarity of images during inspections, while motion blur represents the impact of UAV speed and

vibrations. These enhancements were selectively applied to ensure the model's robustness in facing real-world operational challenges, contributing to the general adaptability of our system across diverse environmental conditions.

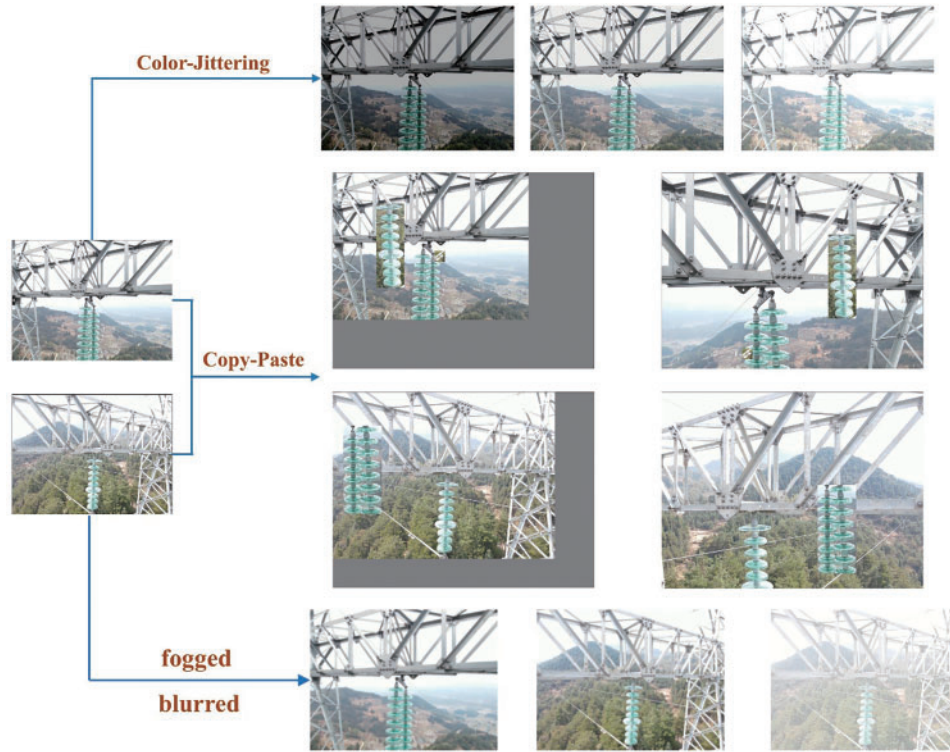


Figure 5: Data augmentation

Ultimately, a dataset comprising 7132 images was generated, encompassing a wide range of diverse environmental scenarios, including mountains, rivers, farmland, grasslands, and cloudy skies. This dataset captures various real-world conditions, such as fluctuating weather patterns, varying lighting levels, and instances of motion blur.

4.2 Evaluation Indicators

In object detection tasks, the evaluation of model performance involves comparing predicted bounding boxes with ground truth (GT) annotations. This comparison yields three fundamental metrics: True Positives (TP), False Positives (FP), and False Negatives (FN). TP counts the number of correctly detected bounding boxes. A bounding box is considered a TP if its IOU with the GT is greater than 0.5, and each GT is matched with only one detection. FP includes those detected bounding boxes that either have an IOU of 0.5 or less with the GT or are redundant detections of the same GT object. FN represents the count of GT objects that the model failed to detect.

These metrics are crucial for calculating two key performance indicators: Precision and Recall. Precision measures the accuracy of the detections and is calculated using Eq. (10):

$$Precision (P) = \frac{TP}{TP + FP} \quad (10)$$

It represents the proportion of true positive detections among all positive detections made by the model. Recall assesses the model's ability to detect all relevant instances and is calculated as Eq. (11):

$$Recall (R) = \frac{TP}{TP + FN} \quad (11)$$

This metric indicates the proportion of actual positives that were correctly identified. However, Precision and Recall alone do not provide a complete picture of a model's effectiveness. Therefore, the Average Precision (AP) metric is introduced. AP represents the area under the Precision-Recall curve for each object category. By averaging these AP values across all categories, we obtain the mean Average Precision (mAP). The mAP, a value ranging from 0 to 1, serves as a comprehensive indicator of the model's overall performance, with higher values indicating superior detection capabilities.

4.3 Comparative Experiments of Backbone Selection

In this study, the AdvResNet50 network module was selected as the feature extraction structure for GFRF R-CNN, with the goal of optimizing its backbone network for improved performance in detecting objects on power lines. To assess the effectiveness of this enhancement, comparative experiments were conducted using different configurations of the Faster R-CNN model, each with a distinct feature extraction module: VGG16, ResNet50, ResNet101, and AdvResNet50.

The dataset for these experiments, comprising images captured by drones, was divided into training, validation, and test sets in a 7:2:1 ratio. After 400 iterations of training, the detection model parameters converged, indicating the models were adequately trained for evaluation.

The results of these comparative experiments, detailed in Table 1, highlight the performance differences among the various configurations: The ResNet, known for their residual units, showed stable performance in detecting power line objects, outperforming the VGG networks. This stability is a key factor in complex object detection scenarios. A slight improvement in accuracy (0.5%) for power line object detection was observed when comparing the Faster R-CNN with ResNet101 to that with ResNet50. This suggests that increased network depth can contribute to better performance, albeit marginally. The most notable improvement was seen with the AdvResNet50. This configuration demonstrated superior information extraction capabilities, especially for small objects like screws. It showed a significant improvement in detection accuracy—4.5% over ResNet50 and 2.1% over ResNet101. Employing the 50-layer AdvResNet as the backbone for the modified Faster R-CNN resulted in improved detection accuracy for all three types of targets. This led to a more stable and reliable overall model performance.

Table 1: Performance of different feature extraction modules (I: Insulators, SH: Shock Hammers, S: Screws)

Backbone	AP(I)	AP(SH)	AP(S)	mAP
VGG16	0.905	0.793	0.647	0.782
ResNet50	0.942	0.852	0.761	0.852
ResNet101	0.936	0.876	0.764	0.857
AdvResNet50	0.949	0.897	0.774	0.873

In summary, the comparison underscores the effectiveness of the AdvResNet50 in enhancing the capabilities of the Faster R-CNN framework. This is particularly evident in complex detection

scenarios involving small and intricate objects, such as those encountered in power line monitoring. The study demonstrates that the choice of feature extraction module significantly impacts the performance of object detection models, with AdvResNet50 emerging as a highly effective option for challenging environments.

4.4 Comparison Experiments

The research conducted a quantitative assessment to evaluate the detection advantages of the proposed GFRF R-CNN, comparing it with four other popular target detection algorithms used in engineering applications: the original Faster R-CNN (ResNet50), SSD, YOLOv3, and YOLOv4. Each of these models, including the GFRF R-CNN with the AdvResNet50 backbone, was trained on a dataset specifically designed for power line target detection. The objective was to perform detection on a unified test set after the models had undergone 400 epochs of iterative training and learning, ensuring that their parameters had adequately converged.

The experimental results, presented in [Table 2](#), aimed to showcase the enhanced performance of the GFRF R-CNN in terms of accuracy, speed, and reliability, especially in detecting complex objects in power line images. The key findings from this comparative analysis can be summarized as follows:

Table 2: Comparison of different methods

Method	mAP@0.5	mAP@0.75	FPS
Faster R-CNN	0.852	0.447	5
YOLOv3	0.793	0.367	14
YOLOv4	0.815	0.419	18
SSD	0.759	0.385	15
YOLOv5	0.822	0.421	20
YOLOv8	0.829	0.425	25
RT-DETR	0.832	0.441	24
Ours	0.893	0.504	8

The GFRF R-CNN continues to demonstrate superior mAP compared to a broader array of models. Specifically, when considering mAP with an IOU threshold of 50%, the improved Faster R-CNN not only outperforms traditional single-stage models such as SSD, YOLOv3, and YOLOv4, but also newer versions like YOLOv5 and YOLOv8, as well as the transformer-based RT-DETR in terms of detection accuracy. With an increased IOU threshold of 75%, the enhancement in the mAP becomes even more pronounced. This further underscores the effectiveness of our method in achieving high accuracy across different operational thresholds, maintaining robust performance against both classical and state-of-the-art models.

While the original Faster R-CNN operated at a speed of 5 Frames Per Second (FPS), the improved version increased this speed to 8 FPS. This enhancement in speed, combined with higher accuracy, underscores the efficiency of the improved model.

[Fig. 6](#) analyzed the training convergence speed of the different models. The GFRF R-CNN model showcased the lowest loss, stabilizing below 0.5. This indicates a more efficient training process and faster convergence compared to the other models.

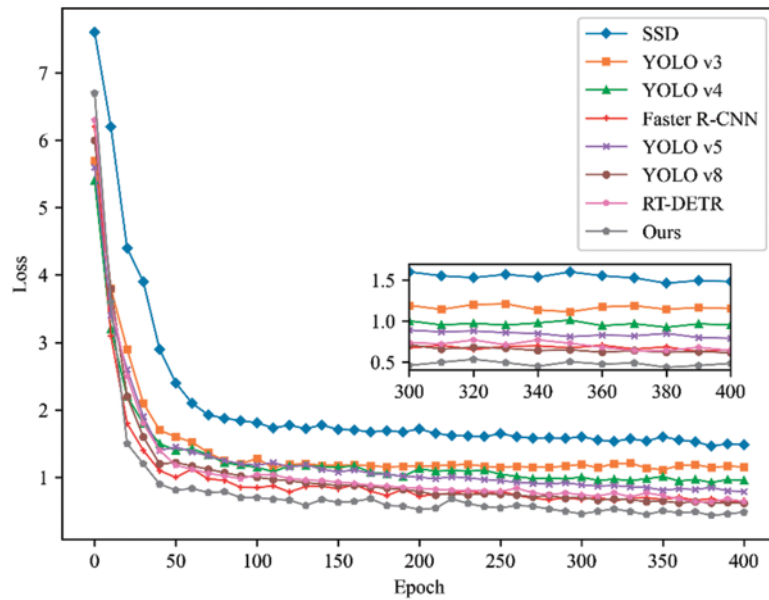


Figure 6: Comparison of loss functions

Overall, the data from this study highlights the efficiency of the improved Faster R-CNN model, particularly in terms of training, convergence, and robustness in complex object detection scenarios. The integration of the AdvResNet50 backbone into the Faster R-CNN framework significantly enhances its capabilities, making it a superior choice for challenging detection tasks such as those involving transmission lines. As depicted in Fig. 7, we describe the Precision-Recall (PR) curve of the GFRF R-CNN.

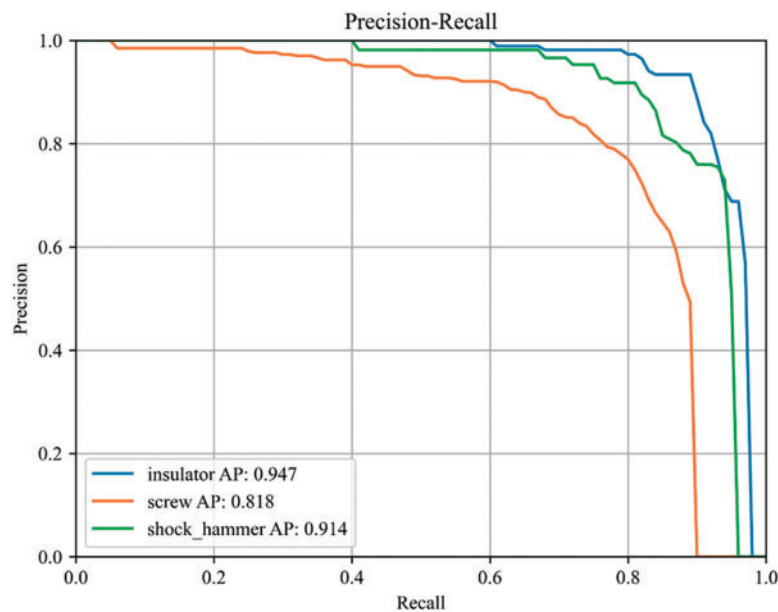


Figure 7: The PR curve of proposed method

In response to the computational complexity and concerns about inference speed, it is imperative to recognize the benefits of utilizing high-resolution images in UAV inspections. Employing images of 8000×6000 pixels, our model captures detailed information over large areas in a single operation, significantly optimizing the efficiency and reducing the logistical challenges associated with UAV flight plans.

However, the use of lower-resolution cameras would require closer and more frequent flights, increasing operational times and energy consumption, ultimately limiting practical application in large-scale deployments.

On the current inference speed, our system operates at 8 FPS with a basic GPU configuration. Although this rate may not suffice for all industrial real-time applications, we foresee a substantial increase in processing speeds with the adoption of high-performance GPUs or advanced edge computing devices, which excel in parallel processing.

4.5 Ablation Experiments and Analysis

The ablation study conducted in this research provides a quantitative assessment of the effectiveness of various components and strategies implemented in the enhanced detection model. By systematically evaluating the impact of different optimization modules, the study illustrates how each component contributes to the overall performance of the detection network. The key findings from the ablation study, as presented in [Table 3](#), are summarized below:

Table 3: Ablation experiments of GFRF R-CNN

ROI-Align	GFR-method	GIOU	AdvResNet50	mAP
	✓	✓		0.863
	✓	✓	✓	0.874
✓		✓	✓	0.859
✓	✓		✓	0.885
✓	✓	✓	✓	0.893

Replacing the traditional backbone with AdvResNet50 led to an increase in the mAP value. This improvement highlights the superior information extraction capabilities of AdvResNet50, confirming its effectiveness in enhancing detection accuracy.

Integrating the GIOU algorithm resulted in an improved mAP. GIOU's ability to consider the non-overlapping areas between the bounding box and the actual object allows for more effective optimization of bounding box regression parameters during training, thereby enhancing the model's precision.

Implementing ROI-Align contributed to an increase in mAP by addressing the dual quantization errors present in the original pooling process. This method of precise feature map extraction is particularly beneficial for detecting small objects, ensuring more accurate localization.

The integration of the GFR method primarily targets the enhancement of the model's detection speed. The ablation study further indicates a notable improvement in detection accuracy. This advancement is credited to the GFR's capability to direct the model's focus specifically to areas where small objects are likely to be located. By doing so, the GFR method effectively increases the precision

of detection, ensuring that the model pays closer attention to regions with a higher probability of containing small, easily missed objects. This strategic focus not only accelerates the detection process but also significantly boosts the overall accuracy of the model in identifying small-scale features.

To further elucidate the impact of our GFR method on the detection of small objects, we conducted an ablation study, the results of which are presented in [Table 4](#). This study was designed to isolate the specific contributions of the GFR method to both detection accuracy and processing speed, providing a clear picture of the trade-offs involved.

Table 4: Ablation experiments of GFR method

Model	mAP (%)	AP_S (%)	Recall (%)	Inference time (ms)	FPS
Baseline	80.8	64.5	70.8	95	10.5
Baseline+P2	85.2	77.4	81.5	198	5.05
Baseline+GFR	89.3	82.9	86.0	123	8.13

The integration of the GFR method significantly improves detection metrics, particularly enhancing the precision of small object detection, with minimal impact on inference time.

The ablation study demonstrates that each optimization strategy contributes uniquely to the overall performance of the detection model. AdvResNet50 improves information extraction, GIOU optimizes bounding box parameters, ROI-Align enhances feature map accuracy, and the GFR algorithm boosts both speed and precision. This comprehensive approach ensures a balanced improvement in both the accuracy and efficiency of the detection model, making it highly effective for complex object detection tasks such as those involving transmission lines. The resolution of the images was 8000×6000 -pixel, and the detection results are shown in [Fig. 8](#).



Figure 8: Detection results

5 Conclusions

This research presents significant enhancements to the Faster R-CNN algorithm, improving its effectiveness in detecting multiple objects within high-resolution images of power lines. The core improvement involves integrating AdvResNet50 as the backbone for feature extraction, which has been rigorously validated through comparative experiments. These experiments demonstrate that the model achieved a mAP of 89.3% at an IOU threshold of 0.5, representing an 11.1% improvement over the baseline Faster R-CNN with VGG16. Additionally, the prediction speed increased from 5 FPS to 8 FPS, underscoring the model's efficiency for real-world applications where both accuracy and speed are critical.

While the primary focus of this study is on power line component detection, the advancements in our model are broadly applicable to other domains involving small object detection in high-resolution images, such as medical imaging, satellite surveillance, and autonomous driving. The improvements in precision and speed make this approach suitable for detecting fine details across various high-resolution imagery contexts, providing potential for wider adoption in other fields where small object detection is essential.

However, the deployment of the proposed model encounters challenges, including the limitations imposed by the data transmission speeds of high-resolution images collected via UAVs. This constraint can significantly impact the real-time processing and analysis capabilities of the system. Additionally, the model's performance is closely tied to the availability of advanced hardware, necessitating high-end computational resources for optimal operation. Acknowledging these challenges, future efforts will explore the development of lightweight, high-resolution object detection models specifically designed for direct deployment on UAVs. This direction aims to mitigate data transmission limitations and reduce reliance on ground-based computational resources, potentially revolutionizing the efficiency and autonomy of UAV-based inspections.

Acknowledgement: The authors are grateful to all the editors and anonymous reviewers for their comments and suggestions and thank all the members who have contributed to this work with us.

Funding Statement: This paper is supported by the Shanghai Science and Technology Innovation Action Plan High-Tech Field Project (Grant No. 22511100601) for the year 2022 and Technology Development Fund for People's Livelihood Research (Research on Transmission Line Deep Foundation Pit Environmental Situation Awareness System Based on Multi-Source Data).

Author Contributions: The authors confirm their contributions to the paper as follows: study conception and design: Xunguang Yan and Wenrui Wang; data collection: Bo Wu and Hongyong Fan; analysis and interpretation of results: Fanglin Lu and Jianfeng Yu; draft manuscript preparation: Xunguang Yan. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Data not available due to confidentiality agreements. The dataset consists of sensitive images from power inspection operations, which cannot be shared to ensure the security and privacy of critical infrastructure.

Ethics Approval: This study did not involve human or animal subjects, and therefore ethical approval was not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

- [1] V. N. Nguyen, R. Jenssen, and D. Roverso, “Intelligent monitoring and inspection of power line components powered by UAVs and deep learning,” *IEEE Power Energy Technol. Syst. J.*, vol. 6, no. 1, pp. 11–21, 2019. doi: [10.1109/JPETS.2018.2881429](https://doi.org/10.1109/JPETS.2018.2881429).
- [2] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017. doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031).
- [3] G. Ghiasi *et al.*, “Simple copy-paste is a strong data augmentation method for instance segmentation,” in *2021 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Nashville, TN, USA, Jun. 20–25, 2021, pp. 2917–2927. doi: [10.1109/CVPR46437.2021.00294](https://doi.org/10.1109/CVPR46437.2021.00294).
- [4] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014. doi: [10.48550/arXiv.1409.1556](https://doi.org/10.48550/arXiv.1409.1556).
- [5] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, “Feature pyramid networks for object detection,” in *2017 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 21–26, 2017, pp. 936–944. doi: [10.1109/CVPR.2017.106](https://doi.org/10.1109/CVPR.2017.106).
- [6] H. Rezaatoughi, 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 *2019 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 15–20. doi: [10.1109/CVPR.2019.00075](https://doi.org/10.1109/CVPR.2019.00075).
- [7] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” in *IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 22–29, 2017, pp. 2980–2988. doi: [10.1109/ICCV.2017.322](https://doi.org/10.1109/ICCV.2017.322).
- [8] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Proc. 2001 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. CVPR 2001*, Kauai, HI, USA, Dec. 8–14, 2001. doi: [10.1109/CVPR.2001.990517](https://doi.org/10.1109/CVPR.2001.990517).
- [9] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, “Object detection with discriminatively trained part-based models,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, 2010. doi: [10.1109/TPAMI.2009.167](https://doi.org/10.1109/TPAMI.2009.167).
- [10] C. Martinez, C. Sampedro, A. Chauhan, and P. Campoy, “Towards autonomous detection and tracking of electric towers for aerial power line inspection,” in *2014 Int. Conf. Unmanned Aircraft Syst. (ICUAS)*, Orlando, FL, USA, May 2014, pp. 27–30. doi: [10.1109/ICUAS.2014.6842267](https://doi.org/10.1109/ICUAS.2014.6842267).
- [11] R. B. Girshick, J. Donahue, T. Darrell, J. Malik, and U. Berkeley, “Rich feature hierarchies for accurate object detection and semantic segmentation,” 2013, *arXiv:1311.2524*.
- [12] R. Girshick, “Fast R-CNN,” in *2015 IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, 7–13 Dec. 2015, pp. 1440–1448. doi: [10.1109/ICCV.2015.169](https://doi.org/10.1109/ICCV.2015.169).
- [13] J. Redmon and A. Farhadi, “YOLO9000: Better, faster, stronger,” in *2017 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 21–26. doi: [10.1109/CVPR.2017.690](https://doi.org/10.1109/CVPR.2017.690).
- [14] J. Redmon and A. Farhadi, “YOLOv3: An incremental improvement,” 2018. doi: [10.48550/arXiv.1804.02767](https://doi.org/10.48550/arXiv.1804.02767).
- [15] A. Bochkovskiy, C. -Y. Wang, and H. -Y. M. Liao, “YOLOv4: Optimal speed and accuracy of object detection,” 2020. doi: [10.48550/arXiv.2004.10934](https://doi.org/10.48550/arXiv.2004.10934).
- [16] W. Liu *et al.*, “SSD: Single shot multibox detector,” in *Computer Vision-ECCV 2016*, Cham, Amsterdam, The Netherlands: Springer International Publishing, 2016, pp. 21–37.
- [17] C. B. Xu, M. Y. Xin, Y. Wang, and J. P. Gao, “An efficient YOLO v3-based method for the detection of transmission line defects,” *Front. Energy Res.*, vol. 11, 2023, Art. no. 1236915. doi: [10.3389/fenrg.2023.1236915](https://doi.org/10.3389/fenrg.2023.1236915).
- [18] X. Liu, Y. Li, F. Shuang, F. Gao, X. Zhou and X. Chen, “ISSD: Improved SSD for insulator and spacer online detection based on UAV system,” *Sensors*, vol. 20, no. 23, 2020, Art. no. 6961. doi: [10.3390/s20236961](https://doi.org/10.3390/s20236961).
- [19] W. Bao, Y. Ren, N. Wang, G. Hu, and X. Yang, “Detection of abnormal vibration dampers on transmission lines in UAV remote sensing images with PMA-YOLO,” *Remote Sens.*, vol. 13, no. 20, 2021, Art. no. 4134. doi: [10.3390/rs13204134](https://doi.org/10.3390/rs13204134).

- [20] W. Zhao, M. Xu, X. Cheng, and Z. Zhao, "An insulator in transmission lines recognition and fault detection model based on improved Faster R-CNN," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–8, 2021. doi: [10.1109/TIM.2021.3112227](https://doi.org/10.1109/TIM.2021.3112227).
- [21] Z. Zhang, H. Ni, M. Liu, Z. Zhang, G. Liu and S. Cheng, "Research on bird nest image recognition and detection technology of transmission lines based on improved Faster-RCNN algorithm," in *2023 5th Asia Energ. Elect. Eng. Symp. (AEEES)*, 23–26 Mar. 2023, pp. 218–222. doi: [10.1109/AEEES56888.2023.10114337](https://doi.org/10.1109/AEEES56888.2023.10114337).
- [22] C. Chen, G. W. Yuan, H. Zhou, and Y. Ma, "Improved YOLOv5s model for key components detection of power transmission lines," *Math. Biosci. Eng.*, vol. 20, no. 5, pp. 7738–7760, 2023. doi: [10.3934/mbe.2023334](https://doi.org/10.3934/mbe.2023334).
- [23] C. Liu, Y. Wu, J. Liu, Z. Sun, and H. Xu, "Insulator faults detection in aerial images from high-voltage transmission lines based on deep learning model," *Appl. Sci.*, vol. 11, no. 10, 2021, Art. no. 4647. doi: [10.3390/app11104647](https://doi.org/10.3390/app11104647).
- [24] Z. Liu, G. Wu, W. He, F. Fan, and X. Ye, "Key target and defect detection of high-voltage power transmission lines with deep learning," *Int. J. Elect. Pow. Energ. Syst.*, vol. 142, no. 3, 2022, Art. no. 108277. doi: [10.1016/j.ijepes.2022.108277](https://doi.org/10.1016/j.ijepes.2022.108277).
- [25] Y. Zhai, X. Yang, Q. Wang, Z. Zhao, and W. Zhao, "Hybrid knowledge R-CNN for transmission line multifitting detection," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021. doi: [10.1109/TIM.2021.3096600](https://doi.org/10.1109/TIM.2021.3096600).
- [26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 27–30. doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [27] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *2017 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 21–26. doi: [10.1109/CVPR.2017.195](https://doi.org/10.1109/CVPR.2017.195).
- [28] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *2018 IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, 18–23 Jun. 2018, pp. 7132–7141. doi: [10.1109/CVPR.2018.00745](https://doi.org/10.1109/CVPR.2018.00745).
- [29] B. Graham and L. van der Maaten, "Submanifold sparse convolutional networks," 2017. doi: [10.48550/arXiv.1706.01307](https://doi.org/10.48550/arXiv.1706.01307).