

Article

Deep-Learning-Based Detection of Transmission Line Insulators

Jian Zhang ¹, Tian Xiao ², Minhong Li ¹ and Yucai Zhou ^{2,*}

¹ School of Design and Art, Changsha University of Science & Technology, Changsha 410114, China; zhangjiancsc@outlook.com (J.Z.); minhongli@outlook.com (M.L.)

² School of Electrical and Information Engineering, Changsha University of Science & Technology, Changsha 410114, China; xiaotian0622@outlook.com

* Correspondence: zycail@163.com

Abstract: At this stage, the inspection of transmission lines is dominated by UAV inspection. Insulators, as essential equipment for transmission line equipment, are susceptible to various factors during UAV detection, and their detection results often lead to leakages and false detection. Combining deep learning detection algorithms with the UAV transmission line inspection system can effectively solve the current sensing problem. To improve the recognition accuracy of insulator detection, the MS-COCO pre-training strategy that combines the FPN module with a cascading R-CNN algorithm based on the ResNeXt-101 network is proposed. The purpose of this paper is to systematically and comprehensively analyze mainstream insulator detection algorithms at the current stage and to verify the effectiveness of the improved Cascade R-CNN X101 model by combining the mAP (mean Average Precision) value and other related evaluation indices. Compared with Faster R-CNN, Retina Net, and other detection algorithms, the model is highly accurate and can effectively deal with the false detection, leakage, and non-recognition of the environment in online special detection. The research in this paper provides a new idea for intelligent fault detection of transmission line insulators and has some reference value for engineering applications.

Keywords: insulator; image processing; deep learning; target identification; neural network



Citation: Zhang, J.; Xiao, T.; Li, M.; Zhou, Y. Deep-Learning-Based Detection of Transmission Line Insulators. *Energies* **2023**, *16*, 5560. <https://doi.org/10.3390/en16145560>

Academic Editors: Ignacio Mauleón, Ke Xing and Bin Huang

Received: 24 May 2023

Revised: 7 July 2023

Accepted: 17 July 2023

Published: 23 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Insulators are hung on power lines, increasing the transmission distance, reducing the current loss, and counteracting some of the capacitive effects of the circuit. When insulators are exposed to the external environment for a long period, they are susceptible to the environmental climate and other factors that can lead to rusting and breakages. To ensure a stable power line operation, insulators in power lines must be inspected regularly to eliminate fault factors early and minimize their impact [1–3]. The maturity of drone inspection technology has greatly improved the efficiency of power line inspection while placing immense strain on the work of inspection. Due to the huge amount of image data collected by drones, the efficiency of manual inspection methods is extremely limited and it is difficult to meet the requirements of the inspection task [4]. Therefore, there is a trend toward the use of image processing technology and machine vision technology to carry out automatic detection of aerial survey data.

Recent years have seen innovation in deep learning methods and rapid improvements in computer hardware performance, and a variety of target detection methods based on deep learning techniques have been proposed sequentially. The application of deep learning methods to the field of electrical power sensing and the design of fault detection algorithms for aerial inspection images using the corresponding techniques are of great importance [5–7]. Here, we take insulators in the transmission line as detection objects and combine the corresponding image processing and annotation methods to construct a new insulator dataset to ameliorate the problem of poor detection accuracy due to an insufficient

sample size. As well as comparing and analyzing the features and differences of the current stage of deep learning algorithms, corresponding methods are used to improve them, and the efficacy of the improved models is tested by a variety of evaluation indices. In the research presented in this paper, we improve the accuracy of the detection algorithm, satisfy the task requirement of accurately identifying insulators in aerial imagery, and provide a new way of thinking to realize smart fault detection of transmission line insulators.

As computer graphics' computing capability has improved, machine vision technology has been extensively used in the fault detection of power line isolators [8–10]. During the inspection process, insulator images are first acquired using acquisition hardware such as drones, followed by the construction of corresponding image-based datasets, and finally, isolators are combined with target detection algorithms for localization as well as fault-type analysis. There are two sorts of insulator fault detection algorithms available at this stage; the first is the algorithm combined with traditional image processing technology, and the second is the algorithm combined with deep learning. At this point, there are still many problems with the traditional manual inspection method, and early research in this area was carried out using image processing techniques for insulator fault detection. The traditional methods focus primarily on isolators with different edge characteristics, such as different texture characteristics, different grayscale characteristics, and different color characteristics, by which the object is compared with the image parameters to be measured and then calibrated based on the isolator data to predict whether the isolator belongs to the fault class [11,12].

Several aspects of deep learning detection methods greatly outperform traditional detection methods and are excellent concerning the effect of feature extraction on the original image. To detect insulator faults, in the literature [13], CNN networks are applied to image feature extraction by fusing the SOM network with the corresponding feature maps and also by combining the superpixel algorithm to aggregate maps of pixels with common visual features as a means of obtaining clearer images of the edges of objects. In the literature [14], the feature extraction network makes use of U-Net, which merges the deep and shallow images in the convolutional layer and then sequentially locates and identifies the isolating objects for both shallow and deep images. This method makes the detection effect enhanced up to a point but is susceptible to the influence of shallow background complications. As a result, the search method in this paper for obtaining excellent results is limited to datasets with clear backgrounds. In the literature [1], the isolation detection algorithm uses SSD, which is first trained using the SSD algorithm to achieve the initial training effect of the training set, and simultaneously begins training the secondary optimization for objects with differing levels of interference and background complexity depending on the weight ratio, which improves the robustness of the detection method as well as the adaptivity of the network model, and this method can efficiently cope with and handle multiple complex sensing environments. In the literature [15,16], they propose a novel insulator detection algorithm with a YOLOv3-based network model, which enhances its diversity on that basis and is endowed with more detector training angles. This allows the detection algorithm to detect isolating objects facing different angles, effectively improving the adaptive ability of the algorithm. They propose an isolation detection algorithm based on the Faster R-CNN network model that has been proposed in the literature [17], including the powerful performance of this model's feature extraction network, deepening the computational depth of the model while at the same time increasing a lower computational pressure. In addition, the algorithm classifies the region of interest and iteratively corrects the prediction framework based on the correction coefficient, which effectively improves the accuracy of the algorithm in insulator identification. It can also merge FCN networks to semantically segment isolator datasets in complex settings.

In summary, it is important to investigate efficient and accurate insulator fault detection algorithms using image processing and deep learning technologies to build a smart power inspection system, determine insulator locations, and detect fault zones. In this paper, we will apply the theory of deep learning to the industrial sensing system framework

from the point of view of deep learning detection methods and propose an MS-COCO pre-training strategy, combined with the FPN module and ResNeXt-101 network to enhance the Cascade R-CNN algorithm, to improve the recognition accuracy of insulator detection algorithms.

2. Processing of Image Datasets

All of the image data used in this paper are obtained from data that are publicly available on Github from the National Grid and transmission line research institutions. There are problems with the data obtained by the above means, such as poor clarity and low capacity, which make the study of image detection algorithms somewhat limited. This is mainly due to the more tedious project of filming a power line inspection, due to the difficulty of conducting and implementing it, as well as data privacy; it is not easy to obtain a large, publicly available dataset of insulator images of the power line environment.

2.1. Image Pre-Processing

When a UAV inspection system acquires imagery, its work is typically outdoors and shot at high altitudes, which is highly susceptible to interference by weather, light, and other factors. The acquired images in this environment tend to be poor, including a lack of clarity, uneven light and darkness, and other issues. This image data were directly used in the subsequent study to train the deep learning detection algorithm, which would largely affect the effectiveness of the model training due to a lack of obvious image characteristics. Before training the detection model, suitable image preprocessing techniques are adopted to improve the best image features as well as to remove interference information from the images. This improves the image quality of the training dataset and allows the detection model to be trained more efficiently.

For the dataset obtained in this paper, there are large differences in light and dark as well as noise pollution due to environmental factors. In this paper, we adopt the corresponding image preprocessing methods for both of these situations.

2.1.1. Image Enhancement

A variety of factors affect the image capture process during drone inspections of transmission lines. If the background of the shot is too bright, if it is a sunny day, or if it is facing the sun, the brightness of the image will be particularly high. If the background of the shot is too dark, or if it is cloudy, the brightness of the image will be too low. If the background is close to the color of the insulator, the isolating features in the image will be less obvious. All of these situations will impact the effect of the detection algorithm. This paper seeks to address this problem by first using the technical means of image enhancement to adjust the contrast of image pixels and increase their luminance.

We choose to use histogram equalization in this paper. To draw the histogram, we rely on the statistical probability of occurrence of different gray values and then maintain a uniform distribution of the number of pixels in the region using a stretch operation, as a means of reducing the bilateral valley contrast and enhancing the top contrast. The histogram equalization of the color images is performed like that of the grayscale maps when the three color channels in the image are processed independently. In the following content, we describe the equalization of the grayscale map histograms.

If the variable r is used to denote the grayscale of the image to be processed and s is used to denote the output grayscale value, the mathematical method of calculating Equation (1) for this process is as follows:

$$s = T(r) \quad (1)$$

where the value domain of T in the mapping function $T(r)$ has to satisfy two conditions ($L = 256$, $T(r)$ lies between 0 and $L - 1$ and r is monotonically increasing on 0 to $L - 1$).

The cumulative distribution function (CDF) accurately meets the above conditions and is often used to express the probability distribution of random variables. Its function is shown in Equation (2).

$$s = T(r) = (L - 1) \int_0^r p_r(w) dw \quad (2)$$

where w is the dummy variable for the integration. The right-hand side of the equation is the cumulative distribution function of the random variable r .

Because the image pixel distribution approximates a discrete function, Equation (2) can again be converted to Equation (3) as follows:

$$s = T(r) = (L - 1) \sum_{i=0}^r p_r(i) \quad (3)$$

where the probability of occurrence of the i -th gray level in the image is represented by $p_r(i)$.

Equation (3) can eventually be written in the form of Equation (4) as follows:

$$s = T(r) = (L - 1) \sum_{i=0}^r \frac{h(i)}{n} \quad (4)$$

where n is the total number of pixels in the image, and $h(i)$ is the number of pixels of each gray level in the histogram.

After the operation of the histogram equalization method for the image, the comparative effect is shown in Figure 1.



Figure 1. Operation of histogram equalization method for images: (a) before adjustment; (b) after adjustment.

2.1.2. Image Filtering

The process of generating, transmitting, and storing image data is susceptible to noise, mainly impulse and Gaussian noise, when the UAV inspection system is acquiring images. For the characteristics of this image dataset, this paper adopts Gaussian filtering and median filtering methods to eliminate the noise in the images.

(1) Median filtering

Median filtering is a non-linear filtering method that is generally used as a method to eliminate impulse noise (pretzel noise). Pixels in the neighborhood are first sorted by their gray value and then the gray value of the central pixel is calculated based on this result. The median filtering method adjusts the window value according to the variation magnitude of the noise to obtain a better filtering effect. Its calculation, Equation (5), is as follows:

$$g(x, y) = \text{median}\{f(x - i, y - j)\} \quad (i, j) \in W \quad (5)$$

where $f(x, y)$ and $g(x, y)$ are the original image and the processed image, respectively, the sliding window is denoted by W and the median is denoted by $\text{median}\{\}$.

(2) Gaussian filtering

Gaussian filtering is a linear filtering method, which is generally used to eliminate Gaussian noise. Gaussian noise has the characteristic that the probability density function is Gaussian distributed. The output of Gaussian filtering takes the value of the weighted average of the pixels in the neighborhood, and since the Gaussian function is single-valued, the closer to the center of the image, the more the pixel weighs. The Gaussian function is rotationally symmetric in two dimensions and has equal smoothness in all directions. Its calculation, Equation (6), is as follows:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (6)$$

where (x, y) is the point coordinate and σ is the standard deviation. The smoothing of the filter as well as the width are determined by the parameter σ . The larger the value of σ , the smoother the image.

The effect after the filtering process is shown in Figure 2. Through the comparative analysis of the effect graphs before and after the processing, it is found that the insulator features in the image are more obvious after the filtering process, which improves the detection effect of the subsequent experimental training model.



Figure 2. Filtering effect: (a) before adjustment; (b) after adjustment.

2.2. Dataset Augmentation

In the samples obtained in this paper, the number of normal insulators is much larger than that of faulty insulators, resulting in a significant difference between the number of positive and negative samples. The unbalanced samples will seriously affect the convergence of the training model and the detection effect, so this paper expands the Matlab database for the insulator data with obvious differences and constructs a dataset suitable for the training model.

In this paper, rotation, translation, and fuzzy processing are used to expand the sample. Since the insulator in the image is large in size and centrally located, a random range of rotation angle parameters and translation parameters is set to prevent the object from falling out of the image after processing. In the rotation processing, the corresponding rotation center point is set, and the image is rotated around the center point to achieve the corresponding effect map. The matrix representation, Equation (7), is as follows:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (7)$$

where the coordinates of each pixel in the original image are represented by (x, y) , and the corresponding coordinates of each pixel in the rotated processed image are represented by (u, v) . The rotation angle is denoted by θ .

The translation moves the original position of the image by a certain distance in four directions: up, down, left, right, and center. Its matrix transformation shape, Equation (8), is as follows:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (8)$$

where the horizontal travel distance is represented by t_x , and the vertical travel distance is represented by t_y .

The image is augmented by the Matlab database and the effect is shown in Figure 3.

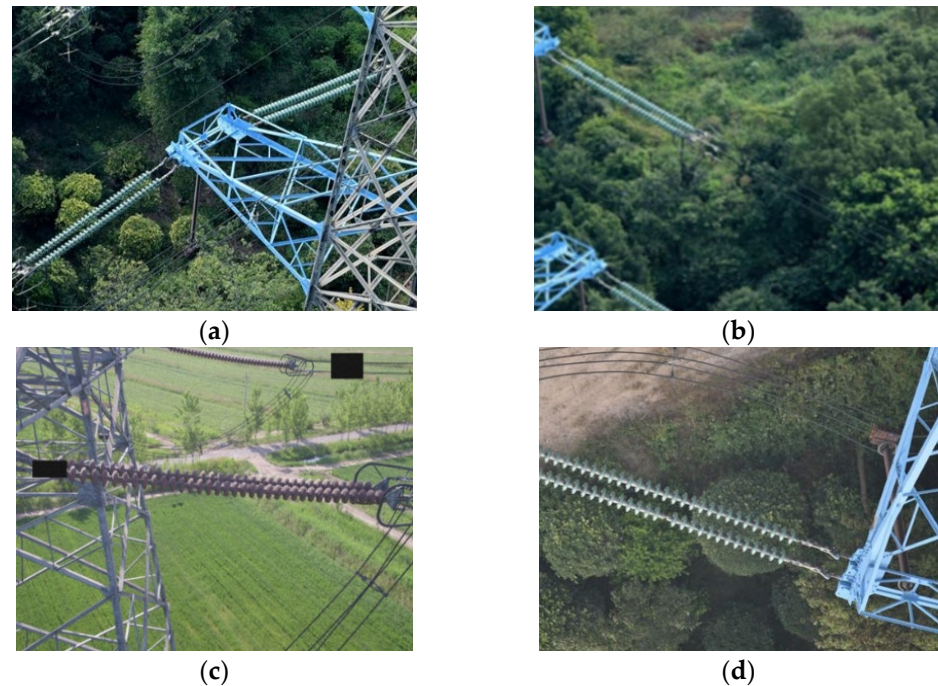


Figure 3. Enlargement effect: (a) flip; (b) kernel ambiguity; (c) add square area; (d) rotate 180°.

To define the experimental dataset: images containing isolators that are normal are used as the positive samples, and images containing insulators with defects are used as the negative samples. There are 852 positive and 118 negative samples in the original dataset, and the amplified samples are filtered to yield 6000 samples. To ensure the reliability of the training model, when training and test set samples are made, the target with large image and feature differences is chosen. Among them, 4800 images were taken for the training set and 1200 for the test set, and the ratio of positive to negative specimens was 1:1 within each dataset.

3. Improved Cascade R-CNN-Based Insulator Detection Algorithm

In this paper, we propose an MS-COCO pre-training strategy to improve the accuracy of the insulator detection algorithm by combining the FPN module and ResNeXt-101 network to improve the Cascade R-CNN algorithm.

3.1. Cascade R-CNN

The Cascade R-CNN algorithm consists of four main modules, including the RPN module (regional proposal network), convolutional neural network module, region of interest pooling (ROI) module, multiple classifiers (Softmax1, Softmax2, and Softmax3), and regressors (B1, B2, and B3). The input image is preprocessed and the features of the image target are extracted in the convolutional layer. Based on the mapping relationship of the features, the candidate frames of the probabilistic presence targets are calculated in the

region generation network. In the ROI pooling module, the feature map is scaled to a fixed size and then sent to the fully connected layer to compute the low-dimensional feature vectors, and the results are output to a detector in the form of a cascade. The structure of the Cascade R-CNN algorithm is shown in Figure 4.

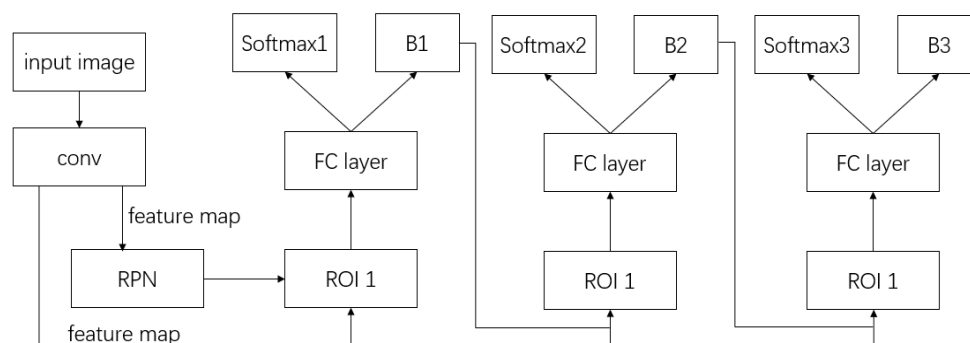


Figure 4. Cascade R-CNN.

The algorithm treats the target as a positive sample and the background as a negative sample. To reduce the difference between the number of positive and negative samples in the high-threshold network and improve the accuracy of the low-threshold network, in each step, the algorithm sets the threshold intersection over union (IOU) for classification and bounding box regression with stepwise augmentation. Except for the first detection module of the algorithm, the input information of the subsequent detection modules is adopted from the output information of the previous detection model. By increasing the number of cascade layers, the IOU threshold is gradually increased and discredited, and the accuracy of localizing the output and classifying the network at each level is gradually improved, and each output is then output to subsequent networks with higher precision. As a result, the Cascade R-CNN algorithm is capable of performing higher-quality detection tasks.

To explore the effect of stage number and IOU values on the experimental results, the AP of COCO 2017 was used for the evaluation in this study, and as can be seen in the table, adding a second stage significantly improved the baseline detector, and adding a third stage also showed a small improvement. There is a small decrease in the AP with the addition of the fourth step, which performs best at high levels of IOU, but the three-step cascade achieves the best compromise between the cost and AP performance.

The Cascade R-CNN algorithm uses three cascade stages for classification and regression, which can provide higher localization accuracy. As can be seen from Table 1, the AP is highest when the IOU thresholds in the cascade stages are set at 0.5, 0.6, and 0.7, respectively.

Table 1. The impact of the number of stages in Cascade R-CNN.

Stages	Test Stage	AP	AP50	AP60	AP70	AP80	AP90
1	1	34.8	57.0	51.9	43.6	29.7	7.1
2	1~2	38.2	57.9	53.6	46.7	34.6	13.6
3	1~3	38.8	57.8	53.4	46.9	35.8	15.8
4	1~3	38.8	57.4	53.2	46.8	36.0	16.0
4	1~4	38.5	57.2	52.8	46.2	35.5	16.3

The IOU threshold of the detection network is set to 0.5, and the anchor frame is input into the network as follows:

- (1) When the IOU between the target frame and the anchor frame is > 0.5 , it is determined that the detection target is included in the anchor frame. The regression loss is

introduced to fine-tune the edge box positions and calculate the initial classification score. After the correction of the regressor, the generated new region is sent to the screening candidate box and finally output to the detection network with an IOU threshold of 0.6.

- (2) When the IOU of the target frame and the anchor frame is > 0.6 , the target is determined to be correctly detected. According to the loss function, the edge frames are adjusted, the regression is corrected for the second time, and the score of the second classification is also calculated. According to this law, the score and position coordinates of the final classification of the target are calculated.

3.2. MS-COCO Pre-Training Strategy

As part of the process of building a network model to perform a specific task of image classification and detection, we initialize the parameters randomly and then train and tune the network until the network's losses are continually reduced. The initialization parameters fluctuate repetitively during model training. Once better results are obtained, information, such as model parameters, is stored so that better results can be obtained the next time that a similar task is performed, a process known as pre-training.

Task-related models (CNNs) for visual detection are typically obtained by training on ImageNet, which has a fairly large dataset with considerable image variety, and it is straightforward to directly apply CNN models to their datasets with corresponding problems. However, the idea of using this data directly to train the network is not feasible when the number of datasets is not sufficient, since the key factor for the efficient detection of deep learning methods is a large number of labeled training sets. It will be difficult for even the best network model to achieve high detection accuracy if only a small training set volume is used. So the pre-trained model must be tuned accordingly.

The experiments in this paper introduce the pre-training strategy of the MS-COCO while adapting the employed deep learning detection algorithm accordingly to its dataset in order to obtain better results.

3.3. FPN Module

The target detection process typically faces the problem of multiscale variation. Many networks at this stage use single high-level features to address this issue. The Faster R-CNN algorithm performs target classification as well as regression processing by downsampling the number of convolutional layers four times. The shortcoming of this processing method is that when the object is a small target, it is easy to lose an object due to little pixel information during downsampling. When there are large differences in the detection objects, algorithms nowadays more commonly use the image pyramid method to improve multiscale variations. This method solves the problems mentioned earlier to some extent but greatly increases the computational effort of the algorithm.

The goal of this paper is to analyze the structure of each deep learning algorithm and introduce the Feature Pyramid Network (FPN) structure to adapt to the presence of targets with multiscale variations during detection. This method not only extracts low-resolution feature maps with strong semantic information but also feature maps with high resolution and low semantic information as well as rich spatial information can be extracted. Figure 5 shows the structure of the feature pyramid.

In semantic segmentation, this structure closely approximates the UNet structure. To achieve a large number of feature layers containing strong semantic information, the feature point downsampling operation is first performed continuously, and the upsampling operation is then performed again to increase the scale of the feature layers and use the feature maps at the largest scale to detect small objects. During this process, it is necessary to stack feature layers with the same scale in both upsampling and downsampling to ensure that the characteristics and information of the small targets are obtained efficiently. Its features are as follows:

- (1) In the feature pyramid, each layer is merged with features from the top layer.

- (2) The top layer of the convolutional network undergoes (1×1) convolution to generate the top layer of the pyramid, while the other layers are sampled over the features in the top pyramid plus the corresponding convolution layer (1×1) , and the pyramidal features from each layer are computed and output to the convolution (3×3) to compute the final features.
- (3) Each layer of pyramid features has a depth of 256 pixels.
- (4) None of the additional convolutions use non-linear activation functions.
- (5) Each layer in the feature pyramid is detected and classified according to its characteristics.
- (6) The convolution layers are related to the feature pyramid by having the same feature size.

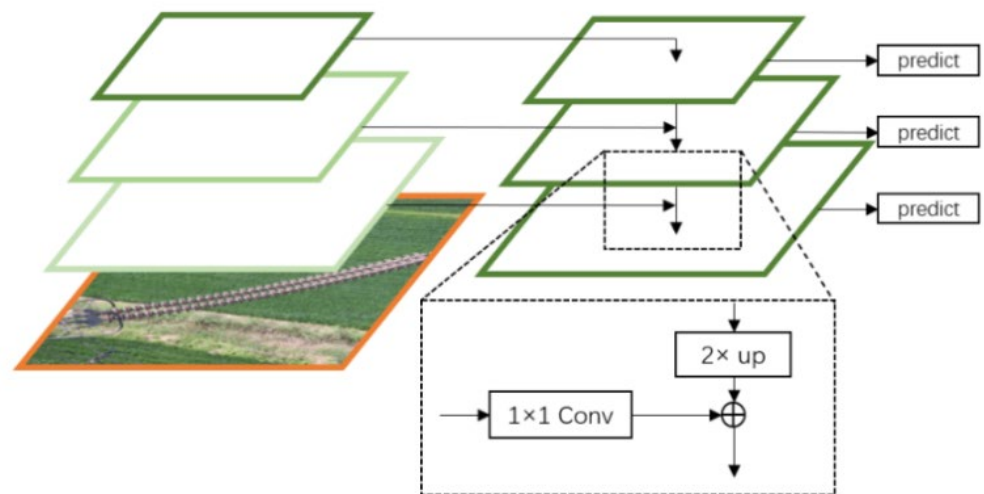


Figure 5. FPN module.

3.4. ResNeXt-101 Network

The ResNeXt-101 network still uses the repetition layer strategy, and the number of paths is increased, based on which a novel split transform fusion strategy is proposed. In this network, modules are correspondingly transformed in the low-dimensional embedding, and all outputs are summed and aggregated while using the same topology for each trajectory. Figure 6 shows the structure of the ResNeXt-101 lattice.

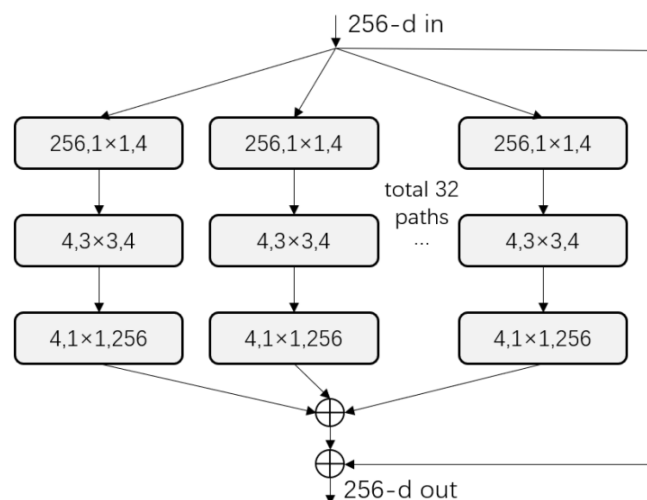


Figure 6. ResNeXt-101 network.

In the above figure, each box represents one layer, where the $(256, 1 \times 1, 4)$ module represents the channel of the input image, the $(4, 3 \times 3, 4)$ module represents the filter size, and the $(4, 1 \times 1, 256)$ module represents the channel of the output data. The path in the structure represents a measurable dimension, but it is different from the width and depth of the channel in the input image. Introducing this measurable dimension can thus effectively improve the accuracy of the detection algorithm when both the width and depth of the objects being detected reduce the training gain of the present model.

The ResNeXt-101 network integrates the advantages of the inception network and the ResNet network. The ResNeXt-101 lattice is equivalent to merging the two models, which can achieve better results by taking advantage of the benefits of each model. These improvements significantly improve the accuracy of the model while only increasing the magnitude of the parameters by a small amount, since there is no difference in the topology and the hyper-parameters are reduced, which facilitates porting of the model.

4. Experimental Setup and Model Training Methods

Firstly, the MS-COCO pre-training strategy is introduced for two-stage models (Faster R-CNN and Cascade R-CNN) and single-stage models (FCOS, Retina Net, and YOLOv7), and an experimental comparison analysis is conducted to verify the effect of the pre-training strategy. The effect of the FPN model is then compared with the Faster R-CNN model before and after the FPN module is equipped to analyze the effect of the FPN model. Lastly, in combination with the ResNet-50, ResNet-101, and ResNeXt-101 backbone networks, respectively, the changes in the loss for the improved algorithm are all registered using the Tensor board to generate the corresponding lossy profile graphs. To determine the effect of training, the same test set is used to test and score the experimental results of constructing different base networks, introducing the FPN module and MS-COCO pre-training strategy for each algorithm, respectively, as well as combining multiple evaluation indices for a validation analysis of the enhanced algorithms.

The training in this experiment is accelerated by CUDA, and the 4000 images are iteratively trained once per cycle, and the cycle is iteratively trained on four GPUs. Firstly, in the training process, the learning rate is set to 2×10^{-2} , and the value is reduced to 0.1 epochs after 8 cycles and to 0.01 epochs after 11 cycles, and the value of the weight decay is set to 0.0005. Since the minibatch value is set to 2, i.e., 2 images are trained on a single GPU, the total number of iterations of the model is $12 \times (4000 \div 2 \div 4) = 6000$ s. During the acquisition of the training samples, each image is sampled 256 times, and the ratio of positive to negative samples is set to 1:1. Then, the branch parameters of the head network and the region generation network are initialized at random, and the minimum value of the IOU of positive samples extracted by the region generation network is set to 0.7 and the maximum value of the IOU of negative samples is set to 0.3 in the process of anchor frame screening. Following experimental screening, if insufficient positive samples are available, the shortage is filled by negative samples. We used group normalization to globally normalize the network parameters.

5. Experimental Results and Analysis

5.1. Environment Environment

The main hardware configurations of the computers used in the experiments of this paper are shown in Table 2.

Table 2. Experiment environment.

Hardware/Software	Parameter
Cpu	Intel Xeon Silver 4210 2.201 GHz
Gpu	GeForce RTX 2080 Ti (11 G)
Ram	64 G
System	Ubuntu 16.04
Language	Python
Deep learning framework	PyTorch

5.2. Analysis of Experimental Results

5.2.1. The Effect of the Improved Module

(1) Introduction of the MS-COCO pre-training strategy

As can be seen in Table 3, the mAP values are higher than those of the original detection algorithm for the three groups of models (Faster R-CNN and ResNet-50), (Faster R-CNN and ResNet-101), and (FCOS and ResNet-50) after the introduction of the MS-COCO pre-training strategy. This indicates that the above problems are improved to some extent after the introduction of the MS-COCO pre-training strategy, and this improvement increases the accuracy of the detection algorithm and has better detection effects. For this reason, the MS-COCO pre-training strategy is introduced in all subsequent detection algorithms to continue the experiments.

Table 3. Introduction of MS-COCO pre-training strategy.

Deep Learning Networks	mAP	mAP_50	mAP_75	mAP_s	mAP_m	mAP_l
* Faster R-CNN and ResNet-50	73.7%	89.9%	83.2%	30.8%	45.3%	77.0%
Faster R-CNN and ResNet-50	76.5%	89.9%	85.9%	47.6%	45.1%	81.5%
* Faster R-CNN and ResNet-101	75.0%	90.7%	84.1%	43.4%	46.1%	79.2%
Faster R-CNN and ResNet-101	76.9%	91.4%	86.6%	47.2%	52.4%	81.8%
* FCOS and ResNet-50	68.7%	88.5%	78.1%	25.7%	40.9%	71.2%
FCOS and ResNet-50	74.5%	91.1%	84.3%	41.2%	44.1%	78.7%

As shown in the table: those with * are the original detection algorithms and those without * are the detection algorithms with the introduction of the pre-training strategy MS-COCO.

(2) Introduction of the FPN module

We can see from Table 4 that the enhanced FPN module for the (Faster R-CNN and ResNet-50) algorithm improves the value of mAP by up to 0.5% over the original algorithm. For medium and large targets, the mAP values do not differ much from the original algorithm, but the mAP value for small detection targets improves by as much as 17.2%. In the original algorithm, target classification and regression processing are performed by downsampling the convolutional layer four times, which can easily result in object losses due to the small amount of pixel information during the downsampling process when the object is a smaller target. In terms of the features of the detection task in this paper, the FPN module is introduced to detect better results, and the module continues to be fed into the subsequent detection algorithm for the experiments.

Table 4. Introduction of FPN module.

Deep Learning Networks	mAP	mAP_50	mAP_75	mAP_s	mAP_m	mAP_l
* Faster R-CNN and ResNet-50	76.0%	91.8%	85.2%	30.4%	51.4%	82.4%
Faster R-CNN and ResNet-50	76.5%	89.9%	85.9%	47.6%	45.1%	81.5%

As can be seen in the table: those with * are the original detection algorithms, and those without * are the detection algorithms after the introduction of the module FPN.

5.2.2. Two-Stage Model Combined to Improve Backbone Network Effectiveness

It can be seen from Table 5 that the mAP values of the Cascade R-CNN algorithm are significantly higher than those of the Faster R-CNN algorithm for the same network as the baseline network. The reason for this is that in the structure of the Faster R-CNN algorithm, only a single R-CNN network is introduced. On the other hand, the Cascade R-CNN algorithm introduces multiple R-CNN networks while cascading them and setting different thresholds for the IOU, which can improve the detection accuracy in a step-by-step manner. Faster R-CNN has a slight advantage over Cascade R-CNN, both in terms of the number of participants and the speedup. The YOLOv7 algorithm has a significant speedup advantage over other algorithms, but is 5.5% less precise than the highest Retina Net in terms of precision. Using the same base network, the FCOS algorithm has a slightly higher mAP value for detecting small-size targets than the other single-stage models when using the ResNet-101 network, but for other sizes, Retina Net has a clear mAP advantage when using the ResNeXt-101 network.

Table 5. Comparison of various models combined with different backbone networks.

Deep Learning Networks	mAP	mAP _50	mAP _75	mAP _s	mAP _m	mAP_l	GFLOPs	M	FPS
Faster R-CNN andResNet-50	76.5%	89.9%	85.9%	47.6%	45.1%	81.5%	206.67	41.13	18.7
Faster R-CNN andResNet-101	76.9%	91.4%	86.6%	47.2%	52.4%	81.8%	282.74	60.13	14.5
Faster R-CNN andResNeXt-101	77.8%	91.9%	88.1%	51.0%	57.8%	84.5%	439.96	98.85	8.7
Cascade R-CNN andResNet-50	76.7%	90.3%	86.2%	46.9%	49.2%	83.4%	234.47	68.93	15.9
Cascade R-CNN andResNet-101	77.5%	90.5%	86.0%	44.0%	49.1%	84.0%	310.54	87.92	12.4
Cascade R-CNN andResNeXt-101	79.9%	91.7%	87.8%	49.7%	54.1%	84.2%	457.76	116.65	8.2
FCOS andResNet-50	74.5%	91.1%	84.3%	41.2%	44.1%	78.7%	196.76	31.84	22.7
FCOS andResNet-101	74.7%	89.7%	83.7%	44.6%	44.6%	79.6%	272.83	50.78	16.6
FCOS andResNeXt-101	75.4%	89.9%	84.6%	50.5%	52.2%	79.9%	434.8	89.61	10.7
Retina Net andResNet-50	75.5%	91.2%	84.7%	42.8%	47.1%	79.0%	205.24	36.15	21.8
Retina Net andResNet-101	76.5%	91.3%	86.0%	43.6%	50.6%	80.5%	281.32	55.14	16.3
Retina Net andResNeXt-101	78.6%	92.8%	87.4%	49.2%	53.6%	82.1%	438.54	93.87	10.4
YOLOv7 andCSPDarknet53	73.1%	92.0%	83.8%	42.6%	45.9%	77.4%	193.89	61.53	51.2

The introduction of the ResNeXt-101 lattice gives the best results with the same model, but at the expense of a larger number of parameters and a slower speedup. For the isolator detection task, however, the mAP value is more important and the image detection accuracy is the kernel, and other metrics can be prioritized in terms of accuracy with a small difference. We see that the Cascade R-CNN in the two-stage model performs better with the introduction of the ResNeXt-101 network, so exploration continues for this model in the following experiments.

5.2.3. Improving the Effect of the Cascade R-CNN X101 Model

In Tables 3 and 4, the Cascade R-CNN with the best detection effect in the two-stage model is compared and analyzed using the Retina Net with the best detection effect in the single-stage model. There are no significant differences in the computation, number of parameters, or speedup between the two, but the former is superior to the latter in terms of detection accuracy. The combination of task requirements and experimental features in this paper thus selects the Cascade R-CNN model for improvement, and the MS-COCO pre-training strategy, the FPN module, and the ResNeXt-101 network are presented.

(1) Enhanced mAP analysis of the model

Figure 7 shows the changes in mAP when the Cascade R-CNN detection algorithm is introduced into the ResNet-50 network, the ResNet-101 network, and the ResNeXt-101 network, respectively.

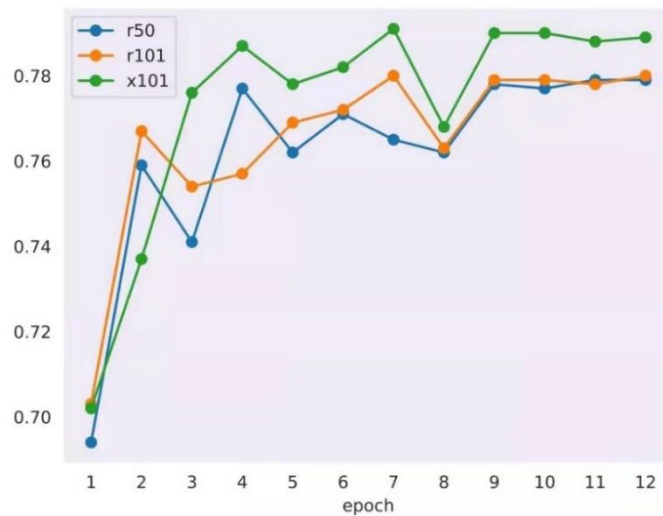


Figure 7. mAP changes.

Comparing the average accuracy, the introduction of the ResNeXt-101 array into the Cascade R-CNN detection algorithm can be found to improve the recognition accuracy of the faulty isolator targets in comparison with the other two arrays. The enhanced model has a significant advantage over other network structures in terms of detection accuracy. The enhanced model is shown to successfully enhance the feature information of the faulty isolator targets in the feature maps generated by the feature extraction network.

(2) Better analysis of the loss curve model

As can be seen by comparing the loss curves of the three networks introduced, ResNet-50, ResNet-101, and ResNeXt-101, in Figure 8, the overall oscillations of the ResNet-101 network are smaller.

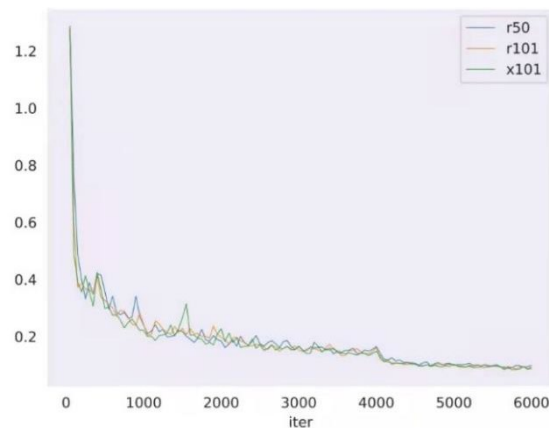


Figure 8. Loss curve.

Comparing the loss curves of the introduction of the ResNet-50, ResNet-101, and ResNeXt-101 networks in Figure 8, it can be seen that the overall oscillation of the ResNet-101 network is lower. By the time the number of samples reaches 2000, the overall fluctuation of the ResNeXt-101 lattice has also weakened considerably, and the loss value for this network is significantly less than the other two networks, which provides evidence that introducing their networks does indeed improve the system stability and convergence.

(3) Analyzing the prediction results from the improved model

Figure 9 shows the comparison of the prediction effect between the improved Cascade R-CNN X101 model and the original model for the same image, as can be seen in Figure 9,

where the left-hand side is the prediction output of the original model and the right-hand side is the improved prediction output.

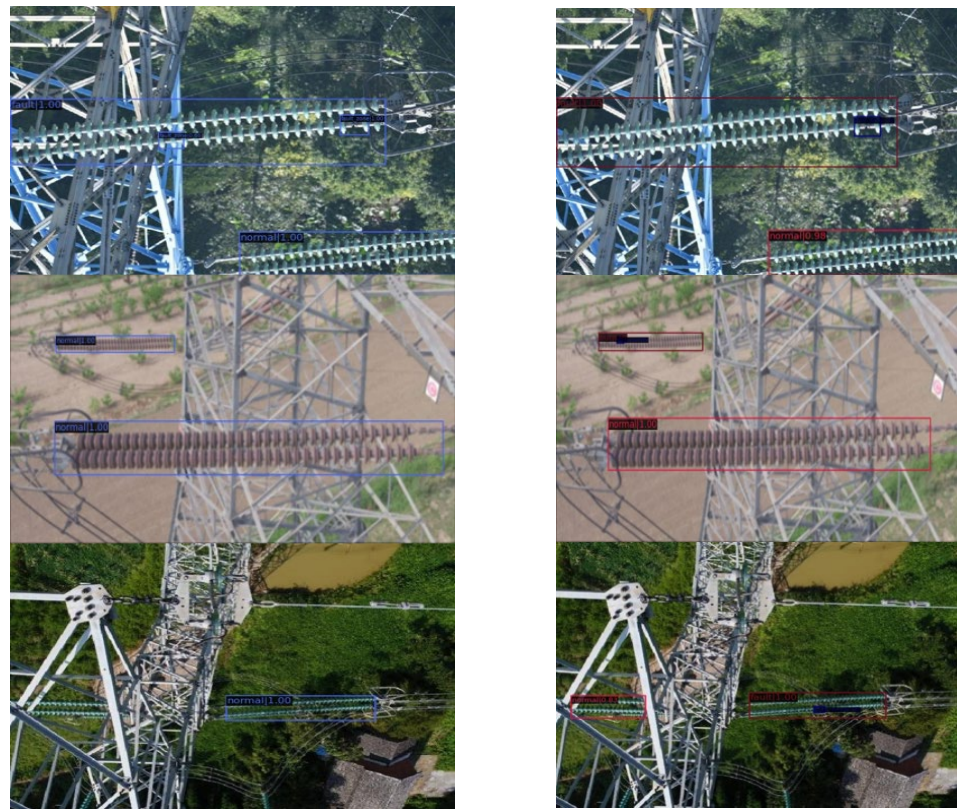


Figure 9. Improved model prediction results.

From comparing the results in the top row of images, the original model can be seen to have undetected the insulator fault detection at the top of the image and undetected the normal insulation part as the failed insulation, which will affect maintenance work efficiency to some degree. Our improved model effectively addresses this problem by accurately detecting the number of defective insulator regions.

As can be seen by comparing the results in the second row of images, note that the original model does not accurately identify the fault zones of the insulators above the figure, and this detection loophole will influence the safety of the power system line components to some degree. In contrast, the improved model accurately predicts all insulator fault zones in the figure.

Comparing the results of the third row of images, we see that the original model is unable to identify isolators on the left-hand side of the figure because they are occluded by obstacles, and leak detection of the isolator will also bring some degree of threat to the safety of the transmission line. However, the enhanced model can still perform recognition detection when the isolator is occluded by an obstacle.

6. Discussion

In this paper, we first describe the development of the detection algorithms in this experiment based on the PyTorch framework as well as the construction of the experimental platform. The network structure is then analyzed for mainstream two-stage detection algorithms (Faster R-CNN and Cascade R-CNN) as well as single-stage detection algorithms (FCOS, Retina Net, and YOLOv7) in the current stage of the research, and the core network compositing system is also being explored according to the pre-training strategies of the MS-COCO, module FPN, and network ResNeXt-101, which are three methods of improvement. The enhancement modules are then introduced for different deep learning

detection algorithms, and the experimental results are compared and analyzed based on the corresponding evaluation indices. Finally, the enhanced Cascade R-CNN X101 model is proposed and the effect of the enhancement is verified. Overall, the enhanced detection algorithm ensures the reliability of the isolator detection and also improves the detection of various special targets. While the improved algorithm increased the number of parameters and increased the elapsed time for detection, this increase in delay does not significantly affect the overall real-time performance of the system and may still satisfy the technical requirements.

For UAV power inspection, this paper studies fault detection of aerial isolator images based on prior research combined with deep learning methods. The rapid development of artificial intelligence, machine vision, and other technologies is leading to the emergence of efficient detection algorithms and network models. The following three aspects can be performed in future research to obtain better results for isolator detection algorithms:

- (1) Expanding the dataset. Due to the great difficulties of obtaining power inspection images, isolators with failures and other types of faults are collected. Obtaining high-definition, high-quality images becomes very challenging due to the presence of the natural environment and other factors. Here, isolator image datasets become rather lacking, resulting in training models with poor accuracy. Future research can investigate more ways in which the image dataset can be extended and improved based on the existing image dataset.
- (2) Efficient network models and algorithms can be further explored in future work. We can investigate the optimal choice of model architecture parameters and continue to simplify the complexity of the sensing network and detection algorithms for the identification of faults such as rusty insulators and cracked insulators can be investigated to achieve a high degree of intelligence in power line detection.
- (3) Enhance the practical value of the model on various platforms and attempt to port the lightweight computational platform into the UAV system to supplement real-time insulator diagnostics by the inspection system, reduce image processing time, and improve task efficiency.

7. Conclusions

China's power system is developing rapidly at this stage, and transmission line coverage is increasing, so the detection effect of the inspection system must be improved to meet the corresponding needs. In this paper, we take transmission line isolators as our research object and propose an enhanced Cascade R-CNN X101 isolator detection algorithm to improve power inspection accuracy more efficiently, given some of the issues that exist in UAV inspection system intelligence. In summary, the work performed in this paper is as follows:

- (1) The collection of aerial images of power line isolators for inspection, the use of corresponding image preprocessing methods, the optimization of resolution, and other image parameters by performing image enhancement and filtering operations without reducing the precision of the algorithm. Based on this, the dataset is expanded using the Matlab database, and then the isolator dataset in COCO format is built using Labelbee software combined with a special annotation system. In a low-quality, low-capacity situation, this technical means effectively improves the detection precision of the training model.
- (2) The structural characteristics of the two-stage detection algorithm, the single-stage detection algorithm, and the backbone network are investigated for the current mainstream insulator fault detection algorithms, and the experimental data are compared and analyzed. In this paper, an MS-COCO pre-training strategy is proposed to improve the Cascade R-CNN algorithm by combining the FPN module and the ResNeXt-101 network, and a matching experimental training method is developed to facilitate verification of improvement.

- (3) In the case of the enhanced Cascade R-CNN X101 network model, experimental simulations are performed by combining several evaluation indices such as mAP value and loss curve, and the efficacy of the enhanced model is effectively verified through comparison and analysis of experimental data with detection algorithms such as Faster R-CNN and Retina Net. The experimental results show that the complexity of the enhanced model is slightly higher than that of the first model, with significantly higher detection accuracies than the other models. When tested on the set of image detection samples, the enhanced model effectively resolves the situation of false detection, missed detection, and unrecognized due to the special environment in patrol detection.

Author Contributions: Conceptualization, J.Z. and T.X.; methodology, Y.Z.; software, T.X.; validation, T.X., J.Z. and Y.Z.; formal analysis, Y.Z.; investigation, J.Z.; resources, T.X.; data curation, M.L.; writing—original draft preparation, J.Z.; writing—review and editing, M.L.; visualization, M.L.; supervision, J.Z.; project administration, M.L.; funding acquisition, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Social Science Foundation of China in the year 2021: grant number 21BC055.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Miao, X.R.; Liu, X.Y.; Chen, J.; Zhuang, S.B.; Fan, J.W.; Jiang, H. Insulator Detection in Aerial Images for Transmission Line Inspection Using Single Shot Multibox Detector. *IEEE Access* **2019**, *7*, 9945–9956. [\[CrossRef\]](#)
2. Du, Y.W.; Wang, H.Y.; Zhao, Y.F.; Sun, L.Q.; Geng, Y.S.; Wang, J.H.; Wang, Z.X. Design of High-Voltage Power Transmission Insulators Based on Ultrasonic Technology. *IEEE Trans. Ind. Electron.* **2023**, *70*, 10740–10749. [\[CrossRef\]](#)
3. Jiang, H.; Zhang, Y.; Lin, J.L.; Zheng, X.G.; Yue, H.; Chen, Y.Z. Developing a new red band-SEVI-blue band (RSB) enhancement method for recognition the extra-high-voltage transmission line corridor in green mountains. *Int. J. Digit. Earth* **2023**, *16*, 806–824. [\[CrossRef\]](#)
4. Fahmani, L.; Garfaf, J.; Boukhdar, K.; Benhadoul, S.; Medromi, H. Unmanned Aerial Vehicles Inspection for Overhead High Voltage Transmission Lines. In Proceedings of the 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco, 19–20 March 2020.
5. Nguyen, V.N.; Jenssen, R.; Roverso, D. Intelligent Monitoring and Inspection of Power Line Components Powered by UAVs and Deep Learning. *IEEE Power Energy Technol. Syst. J.* **2019**, *6*, 11–21. [\[CrossRef\]](#)
6. Song, Z.W.; Huang, X.B.; Ji, C.; Zhang, Y. Intelligent Identification Method of Hydrophobic Grade of Composite Insulator Based on Efficient GA-YOLO Former Network. *IEEE Trans. Electr. Electron. Eng.* **2023**, *18*, 1160–1175. [\[CrossRef\]](#)
7. Yang, Y.L.; Wang, X.L. Insulator detection using small samples based on YOLOv5 in natural background. *Multimed. Tools Appl.* **2023**. [\[CrossRef\]](#)
8. Tao, X.; Zhang, D.; Wang, Z.; Liu, X.; Zhang, H.; Xu, D. Detection of Power Line Insulator Defects Using Aerial Images Analyzed With Convolutional Neural Networks. *IEEE Trans. Syst. Man Cybern. Syst.* **2020**, *50*, 1486–1498. [\[CrossRef\]](#)
9. Chen, W.H.; Cai-Lin, L.L.; Yuan, B.; Jiang, X.B. Effective method to locate self-explosion defects of insulators. *Comput. Eng. Des.* **2019**, *40*, 2346–2352.
10. Zheng, J.F.; Wu, H.; Zhang, H.; Wang, Z.Q.; Xu, W.Y. Insulator-Defect Detection Algorithm Based on Improved YOLOv7. *Sensors* **2022**, *22*, 23. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Nguyen, V.N.; Jenssen, R.; Roverso, D. Automatic autonomous vision-based power line inspection: A review of current status and the potential role of deep learning. *Int. J. Electr. Power Energy Syst.* **2018**, *99*, 107–120. [\[CrossRef\]](#)
12. Peng, X.; Qian, J.; Wang, K.; Mai, X.; Lin, Y.I.; Rao, Z. Multi-sensor Full-automatic Inspection System for Large Unmanned Helicopter and Its Application in 500 kV Lines. *Guangdong Electr. Power* **2016**, *29*, 8–15.
13. Yan, B.; Chen, Q.; Ye, R.; Zhou, X. Insulator detection and recognition of explosion based on convolutional neural networks. *Int. J. Wavelets Multiresolut. Inf. Process.* **2018**, *17*, 1940008. [\[CrossRef\]](#)
14. Sampedro, C.; Rodriguez-Vazquez, J.; Rodriguez-Ramos, A.; Carrio, A.; Campoy, P. Deep Learning-Based System for Automatic Recognition and Diagnosis of Electrical Insulator Strings. *IEEE Access* **2019**, *7*, 101283–101308. [\[CrossRef\]](#)
15. Zhang, Y.; Zhu, X.; Bu, Y.; Ding, W.; Lu, Y. *Detection System of Truck Blind Area Based on YOLOv3*; Springer: Singapore, 2022.

16. Chen, S.; Su, C.; Kuang, Z.; Ye, O.; Gong, X. Real-time detection of UAV detection image of power line insulator bursting based on YOLOV3. *J. Phys. Conf. Ser.* **2020**, *1544*, 012117. [[CrossRef](#)]
17. Ni, H.; Wang, M.; Zhao, L. An improved Faster R-CNN for defect recognition of key components of transmission line. *Math. Biosci. Eng. MBE* **2021**, *18*, 4679–4695. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.