



## OPEN ACCESS

EDITED BY  
Yiyi Zhang,  
Guangxi University, China

REVIEWED BY  
Hao Jiang,  
Fuzhou University, China  
Runsheng Tang,  
Yunnan Normal University, China

\*CORRESPONDENCE  
Zhaoyun Zhang,  
18927491998@163.com

SPECIALTY SECTION  
This article was submitted  
to Smart Grids,  
a section of the journal  
Frontiers in Energy Research

RECEIVED 04 April 2022  
ACCEPTED 30 August 2022  
PUBLISHED 23 September 2022

CITATION  
Zhang Z, Cai D and Zhang Z (2022),  
Review on online operation insulator  
identification and fault diagnosis based  
on UAV patrol images and deep  
learning algorithms.  
*Front. Energy Res.* 10:912453.  
doi: 10.3389/fenrg.2022.912453

COPYRIGHT  
© 2022 Zhang, Cai and Zhang. This is an  
open-access article distributed under  
the terms of the [Creative Commons  
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,  
distribution or reproduction in other  
forums is permitted, provided the  
original author(s) and the copyright  
owner(s) are credited and that the  
original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution  
or reproduction is permitted which does  
not comply with these terms.

# Review on online operation insulator identification and fault diagnosis based on UAV patrol images and deep learning algorithms

Zhaoyun Zhang<sup>1\*</sup>, Delong Cai<sup>2</sup> and Zhi Zhang<sup>1</sup>

<sup>1</sup>School of Computer Science and Technology and School of Electrical Engineering and Intelligencization, Dongguan University of Technology, Dongguan, China, <sup>2</sup>School of Computer Science and Technology, Dongguan University of Technology, Dongguan, China

Artificial intelligence has great potential for use in smart grids. Power system image recognition based on artificial intelligence is an important research direction. The insulator is essential equipment for the power grid and is related to operational safety. Online operating insulator location identification and fault diagnosis technologies based on unmanned aerial vehicle (UAV) patrol the images, and deep learning algorithms have been continuously suggested and developed. These technologies have achieved good results in practical application. By compiling the recent literature on insulator detection technology, three common application scenarios and research difficulties are uncovered: The need for increased detection accuracy and real-time speed; faulty image recognition of complex backgrounds and target occlusion; and multiscale object and small object detection improvements. At the same time, the improved algorithms in the literature are comprehensively summarized, and the performance evaluation indices of various algorithms are compared.

## KEYWORDS

deep learning, insulator, localization recognition, fault diagnosis, complex background, small objects

## 1 Introduction

With the large-scale construction of ultrahigh voltage and the “nine crosses and nine straight” plan, China’s power industry has gradually moved to the forefront of the world. At the same time, the advancements also pose higher challenges to the safe maintenance of the power systems. Currently, most of the inspection work of power lines relies on traditional manual methods, with labour that is not only cumbersome and inefficient but that is also constrained by geographical location. In recent years, UAV patrol technology has been a hot topic in intelligent power grids, and its application to power inspection is also a significant trend in smart power grids (Tong, 2022).

As an indispensable insulating element in power transmission lines, insulators’ operating conditions directly affect the power grid’s reliability and safety. According to statistics, insulator defects cause the highest percentage of the current accidents in power system failures.

Therefore, it is vital to monitor insulator conditions and promptly diagnose faults. By analysing and processing the images captured by inspection robots or UAVs, the fast and accurate detection of insulators in aerial images has become a current hot research topic (Ma, 2021). The UAV captures many high-definition insulator images, but the images show complex insulator backgrounds and varying scales. It would be time-consuming to rely solely on the human processing of image information to identify whether the insulators are in an intact condition. Hence, the implementation of automated processing of image data information and the accurate detection of insulators' status is the key to improving the efficiency of power inspections. Many researchers have devised various methods for location recognition of aerial insulator images. Different convolutional network models are emerging with the rapid development of deep learning. The autonomous learning of features in images through multilayer convolutions not only improves a model's generalization capabilities but also reduces the data dimensionality and redundant information, reducing the computational cost and substantially improving the speed and precision of detection (Zou, 2022).

In the last two decades, object detection has undergone two development periods (Zou et al., 2019), the traditional object detection period and the deep learning-based object detection period. After AlexNet (Krizhevsky et al., 2017) made a splash at the ILSVRC challenge (Jia et al., 2009) in 2012, the application of traditional target detection algorithms decreased, and deep learning gradually became mainstream. This paper summarizes the three common difficulties encountered in insulator detection based on deep learning object detection and the algorithm improvements that were made.

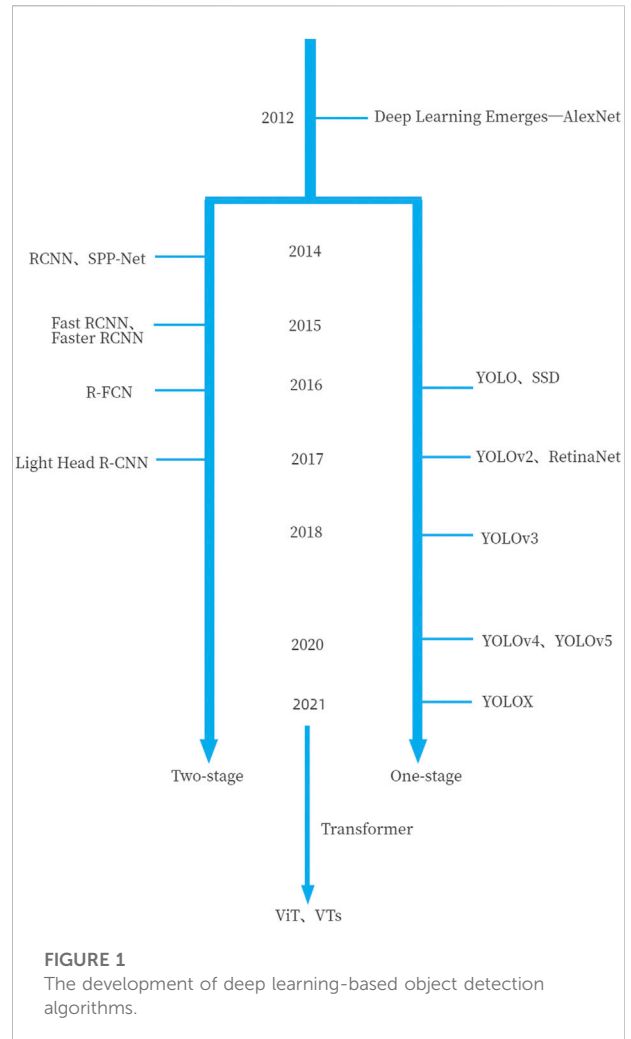
## 2 The deep learning-based object detection algorithm

In the era of deep learning, target detection algorithms include "two-stage detection," "single-stage detection," and transformer structure-based algorithms. Figure 1 shows the development history.

### 2.1 Two-stage detection algorithm

#### 2.1.1 Region-based convolutional neural network

In 2014, Girshick et al. (2013) proposed the R-CNN algorithm. R-CNN changed the previous best-performing but the more complex integrated system and adopted a simple and scalable detection algorithm, which improved the mean average precision (mAP) by more than 30% compared to the previous best results of VOC 2012. Although the R-CNN algorithm has made significant progress compared with previous algorithms, it requires iterative calculations on many overlapping candidate regions, resulting in a prolonged detection speed.



#### 2.1.2 Spatial pyramid pooling network

Later, in the same year, He et al. (2015) proposed the SPP-Net algorithm to address the drawback of repeatedly computing convolutional features in the R-CNN algorithm. SPP-Net generates fixed-length feature layers by employing an alternative pooling strategy and spatial pyramidal pooling, without considering the size of the input image. Additionally, for object deformation, pyramid pooling is highly robust. In addition, pyramid pooling is robust against object deformation. The SPP-Net algorithm effectively improves the detection speed, but there are still some problems. The training of SPP-Net remains multi-stage, and the researchers only fine-tune its fully connected layers while ignoring all previous layers.

#### 2.1.3 Fast region-based convolutional neural network

To address the problems of SPP-Net, in 2015, Girshick (2015) proposed the Fast R-CNN algorithm, which simplified the SPP layer and designed a single-scale region of interest (RoI) pooling layer structure. Introducing a multitask loss function

improves the precision and speed of Fast R-CNN. While Fast R-CNN successfully combines the advantages of R-CNN and SPP-Net, the network (Huang and Zhang, 2020) still uses a relatively time-consuming selective search method to generate candidate regions. It requires the acquisition of RoIs before the images are fed into the CNN network. Therefore Fast R-CNN network is not an end-to-end network in real implementations or actual applications.

#### 2.1.4 Faster region-based convolutional neural network

Faced with these problems, researchers began to focus on using CNNs directly to generate RoIs. In 2015, Ren et al. (2017) proposed the Faster R-CNN algorithm, which introduced the region proposal network (RPN) to generate candidate regions. The RPN shares the full image convolutional network with the detection network, and the two merge into a single network to achieve almost real-time detection. However, the RoI pooling layer between RPN and Fast R-CNN converts the multiscale feature mapping into fixed-scale feature mapping, which directly destroys the translation invariance of the network and is detrimental to the classification of image objects.

#### 2.1.5 Region-based fully convolutional network

In 2016, Dai et al. (2016) proposed region-based fully convolutional networks (R-FCNs) to address the shortcomings of Faster R-CNN. The proposed Position-Sensitive Score Map addresses the difficulties between translation invariance in image classification and the translation variance in target detection. The R-FCN adopts a residual network (He et al., 2016) as a fully convolutional image classifier backbone. It uses a position-sensitive mapping map RoI pool to coordinate the translation invariance and sensitivity across the convolutional layers, improving object localization.

#### 2.1.6 Light head region-based convolutional neural network

Although the detection speed of the R-FCN has been better than that of the R-CNN family of algorithms in the two-stage detection algorithm, the overall architecture design is excessively bulky. To simplify the network and improve the detection speed, in 2017, Li et al. (2017) proposed a new two-stage detector, the Light-Head RCNN. It achieves a good balance of speed and precision by using a thin feature map and a low-cost R-CNN subnet (pooling and a single fully connected layer) that makes the head of the network as light as possible.

## 2.2 One-stage detection algorithm

The one-stage detection algorithm is different from the two-stage detection algorithm. It adopts the idea of regression analysis, which does not need to generate candidate regions and directly obtain the object classification and location

information. It is mainly represented by the YOLO (You Only Look Once) series algorithm and SSD (Single Shot MultiBox Detector) algorithm.

#### 2.2.1 You only look once series algorithm

In 2016, Redmon et al. (2016) proposed the first one-stage network, YOLOv1, to address the common problem of the poor real-time performance of two-stage detection algorithms. Its use of object detection as a regression problem allows for a direct end-to-end optimization of detection performance since the entire detection pipeline is a single network. However, YOLOv1 is ineffective in detecting small objects and is prone to missed detections in cases with object overlaps and occlusions. Therefore, in 2017, Redmon and Farhadi (2017) proposed an improved version of YOLOv1, YOLOv2, which uses DarkNet-19 as the backbone network and introduces improvements such as batch normalized preprocessing, a multiscale training mechanism, and a binary cross-entropy loss function, resulting in a significant increase in recall and precision. To improve the network sensitivity to small object detection, in 2018, Redmon and Farhadi (2018) proposed the YOLOv3 algorithm. It uses Darknet-53 as the backbone network, deepens the number of network layers, and introduces the cross-layer summation operation in Resnet. YOLOv3 is fast in practical applications, while the background false detection rate is low.

In 2020, Bochkovskiy et al. (2020) proposed the YOLOv4 algorithm. Combining many previous research techniques and a wide range of experiments, they investigated the impact of many generalized algorithms on network performance. They found the optimal combination, especially the mosaic augmentation technique, which can effectively solve the “small object detection difficulty” problem. Compared with YOLOv3, YOLOv4 has dramatically improved the overall detection performance and the detection performance for obscured objects. In the same year, researcher Glenn Jocher released an open-source implementation of YOLOv5 and its derivative version on GitHub, whose performance is comparable to YOLOv4. Unlike YOLOv4, YOLOv5 builds on the PyTorch implementation, with simpler support, easier deployment, fewer model parameters and support for conversion to ONNX and CoreML. It is convenient for users to deploy on mobile, embedded devices, etc.

In 2021, a new generation version of the YOLO algorithm, YOLOX, was proposed by MEGVII (Ge et al., 2021). It switches the YOLO detector to an anchorless approach with other advanced detection techniques, namely, the decoupling header and the leading label assignment strategy-SimOTA. YOLOX achieves state-of-the-art results in an extensive range of model comparisons while improving support for the ONNX, TensorRT, NCNN, and Openvino deployment versions with strong industrial applicability.

#### 2.2.2 Single shot multiBox detector algorithm

In 2016, Liu et al. (2016) combined the anchor mechanism of R-CNN and the regression idea of YOLO to propose the SSD algorithm. SSD is the second single-stage detection algorithm in

the era of deep learning, which introduces a multiscale detection technique that performs detection on the feature maps extracted at each scale. SSD has a higher precision for processing smaller images than the other one-stage detection algorithms.

### 2.2.3 RetinaNet

In 2017, [Lin et al. \(2017\)](#) found that the extreme foreground-background class imbalance encountered during dense detector training is the main reason why the precision of the one-stage algorithms lags behind that of the two-stage algorithms. In this regard, RetinaNet was proposed, which introduces a new loss function, “focal loss,” which enables the detector to focus more on difficult samples during training by reconstructing the standard cross-entropy loss. The focal loss allows the one-stage detector to achieve comparable precision to the two-stage detector while maintaining a very high detection speed.

## 2.3 Transformer structure

In 2017, Transformer, the Self-Attention Architecture, was proposed, mainly for natural language processing, and expanded to various domains. In 2020, [Dosovitskiy et al. \(2020\)](#) pioneered the direct application of Transformer to vision, proposing Vision Transformer (ViT). ViT segments an image into multiple nonoverlapping patches, similar to tokens (words) in natural language processing, and performs a series of linear embedding operations on each patch as input to the transformer. However, it is found through experiments that the transformer lacks the inductive biases inherent to CNNs, such as translational invariance and localization. Consequently, the generalization is not good when the training set is insufficient.

In 2021, the second generation of Visual Transformers (VTs) architecture ([Liu et al., 2021a](#); [Yuan et al., 2021a](#); [Yuan et al., 2021b](#); [Hudson and Zitnick, 2021](#); [Li et al., 2021](#); [Wu et al., 2021](#); [Xu et al., 2021](#)) based on ViT was proposed. These second-generation VT networks usually mix convolutional layers with attention layers to provide a local inductive bias for VTs. These hybrid structures have the advantages of both paradigms, the attention layer models global dependencies, while the convolution operation can emphasize local features of the image content. Most of the experiments in this work show that these second-generation VTs outperform the training on ImageNet versus a ResNet of similar size. However, the training results of these second-generation VT networks on small and medium-sized datasets are still unclear.

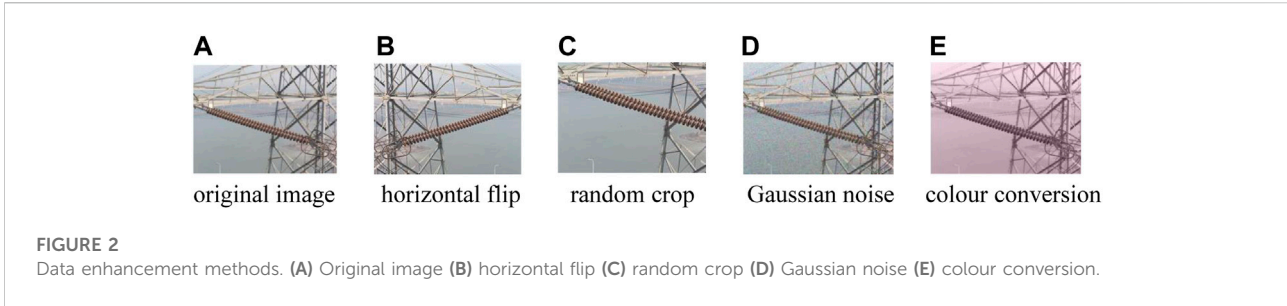
To investigate the training performance of VTs on small and medium-sized datasets later in the same year, [Liu et al. \(2021b\)](#) proposed using other self-supervised tasks and corresponding dense relative localization loss functions. These improvements can increase the accuracy of VTs to a great extent. Especially when training from scratch on small datasets, VTs can significantly improve transformer performance, with the accuracy values rising to 45%.

## 3 Insulator dataset and performance evaluation

### 3.1 Related literature insulator dataset source and production

The current datasets publicly applied to deep learning models do not contain insulator images of the transmission lines, and there is no specialized dataset for aerial insulator image recognition for transmission lines at home and abroad. The datasets that are now publicly accessible are the sample data from GitHub's Insulator DataSet-Chinese Power Line Insulator Dataset (CPLID, GitHub—InsulatorData/InsulatorDataSet: Provide normal insulator images captured by UAVs and synthetic defective insulator images), as well as datasets created by different research institutions on their own. The number of acquired insulator images is limited. In addition, too few data tend to cause an overfitting of the model obtained from training, resulting in the network's inability to learn the true distribution of the data and largely reducing the model's generalization ability.

One of the most straightforward ways to deal with the lack of training data ([Nguyen et al., 2018](#)) is to manually create training data. However, this is a very slow, tedious and expensive process. One possible way to speed up the process is to use a pretrained model and fine-tune it with a small amount of manually created training data to generate more data automatically. When only a small amount of training data is available, experimenters can use data augmentation techniques to improve the training performance. Some simple data enhancement techniques help train deep learning models, such as mirror flipping, random cropping, colour conversion, or adding random noise to random sample images, as shown in [Figure 2 \(Wang, 2021\)](#). Another way to resolve the lack of training data is to use synthetic images. However, effectively combining synthetic and real images in the training of deep learning models is still challenging. Therefore, researchers have applied supervised domain adaptation ([Csurka, 2017](#)) to address this challenge. In this approach, the model is first trained only for synthetic images and images of the task of interest; then, fine-tuning for target tasks that typically have few training examples. In the absence of available training examples, unsupervised domain adaptation ([Ganin and Lempitsky, 2014](#); [Cai et al., 2018](#); [Long and Wang, 2015](#); [Sener et al., 2016](#)) is a potential solution that can adapt a model trained only on synthetic images and images from related tasks for the target task. The class imbalance problem can be solved to some extent by synthesizing images, e.g., normal insulator images with positive samples and faulty insulator images with negative samples. At the same time, in realistic scenarios, there are more faulty insulator samples than normal samples, which results in a positive and negative sample imbalance. A balance of the unbalanced classes is attained by generating more synthetic images for classes with fewer training samples. An alternative solution is the median frequency balancing method, in which



classes with fewer training samples are assigned higher weights during training (Eigen and Fergus, 2014; Kampffmeyer et al., 2016; Badrinarayanan et al., 2017).

Given the identifiability of deep learning models, the dataset constructed in this comprehensive literature follows the library construction method of the public PASCAL VOC dataset. The data annotation tool used for data annotation of the insulator samples is LableImg, while strictly following the PASCAL VOC data format: 1) The annotation frame fits the insulator object tightly; 2) the fuzzy insulators are ignored. LableImg can record the category names and the location pieces of information of the objects in the images and store these pieces of information in the corresponding XML format files for the training of the models.

### 3.2 Detection object and classification

The insulator images based on UAV patrol contain several common defects and corresponding severity levels, as shown in Table 1. The severity level of the insulators includes three kinds of defects: emergency, major and general. The defects listed in Table 1 are essential for deep learning insulator fault diagnosis.

### 3.3 Algorithm performance evaluation metrics

The accuracy measures for insulator identification or fault diagnosis are average precision (AP), mean average precision (mAP), recall (R), and frame per second (FPS). TP indicates that the actual positive sample is predicted to be a positive sample as well, TN indicates that the predicted negative sample is a negative sample, FN indicates that the actual positive sample is predicted to be a negative sample, and FP indicates that the actual negative sample is predicted to be a positive sample.

- (1) Precision, for the predicted results, indicates the proportion of the actual positive samples among the predicted positive samples. The calculation formula is shown below:

$$P = \frac{TP}{TP + FP} \quad (3 - 3 - 1)$$

- (2) Recall, for the original sample, indicates how many positive samples in the sample were correctly predicted. The calculation formula is shown below:

TABLE 1 Insulator defects and severity level.

Defect types	Emergency defect	Major defect	General defect
• Insulator string pin off or insulator string off.	√		
• The number of zero value or broken vases on a string of insulators reaches 1 piece at 35 kV, 3 pieces at 110 kV, 4 pieces at 220 kV, 4 pieces at 500 kV and above.	√		
• Needle insulator string and porcelain cross-arms are severely damaged, and nuts fall off.	√		
• High-voltage pillar insulator with cracks, glaze broken area greater than 1 cm <sup>2</sup> .	√		
• Insulator string cotter pin dislodged.		√	
• The number of zero value or broken vases on a string of insulators reaches 1 piece at 10 kV and below, 2 pieces at 35 kV, 3 pieces at 110 kV, 3 pieces at 220 kV, 3 pieces at 500 kV and above.		√	
• Ground line dangling insulator string along the line tilt angle less than 30° greater than 30°.		√	
• Insulator surface staining and flashing serious or rusting seriously.		√	
• The insulator cotter pin is not open.			√
• Insulators indicate dirt and scale accumulation.			√
• Ground line overhang insulator string along the line tilt angle is less than 30°.			√

$$R = \frac{TP}{TP + FN} \quad (3-3-2)$$

(3) Precision and recall are two contradictory measures. According to the correlation between precision and recall, when precision increases, recall decreases accordingly, and vice versa. Therefore, it is necessary to combine these two parameters and use the AP value to measure the algorithm performance. The calculation formula is shown below:

$$AP = \sum_{n=1}^N P(n) \Delta R(n) \quad (3-3-3)$$

where  $N$  denotes the number of samples in the test set,  $P(n)$  denotes the precision  $P$  when identifying  $n$  samples, and  $\Delta R(n)$  denotes the change in recall  $R$  when the number of identified samples changes from  $n-1$  to  $n$ . For multi-classification ( $K$ -class) detection tasks, the model is usually evaluated using the mean mAP. The calculation formula is shown below.

$$mAP = \frac{AP}{K} \quad (3-3-4)$$

(4) Speed evaluation. An important measure of the detection speed of the detection algorithm is the frame rate per second (FPS), i.e., the number of images that can be processed per second; the larger the FPS, the faster the detection speed and the better the real-time performance of the algorithm. The calculation formula is shown below:

$$FPS = 1 / \left( \frac{T}{F} \right) \quad (3-3-5)$$

where  $T$  denotes the total image processing time consumed by the algorithm and  $F$  denotes the number of image frames processed by the algorithm.

The literature (Chen et al., 2018; Zhao et al., 2019a; Zhao et al., 2019b; Lai et al., 2019; Miao et al., 2019; Pan, 2019; Wu et al., 2019; Yang, 2019; Zuo et al., 2019; Chen et al., 2020; Chen and Min, 2020; Ding, 2020; Ji, 2020; Ji et al., 2020; Pan et al., 2020; Shen et al., 2020; Wu, 2020; Yan and Chen, 2020; Yao and Qin, 2020; Zhou et al., 2020; Gao et al., 2021b; Chen et al., 2021; Liu et al., 2021c; Liu et al., 2021d; Liu et al., 2021e; Huang et al., 2021; Tan, 2021; Tang et al., 2021; Tian et al., 2021; Wang et al., 2021; Yan and Wang, 2021; Yi et al., 2021; Zhang et al., 2021; Zhang and Guo, 2021; Zhu et al., 2021; Gao and Wang, 2022; Huang et al., 2022; Jiang et al., 2022) has used average precision (AP), mean average precision (mAP), recall (R) and frames per second (FPS) metrics to evaluate the improved algorithms. Among them, 26 papers adopted the most frequently used AP metric, 14 papers adopted the mAP metric, 17 papers adopted the R metric and 17 papers adopted the FPS metric. In the insulator detection task, the average precision (AP) is undoubtedly necessary because it is a direct measure of whether the insulators in the aerial images can be correctly identified as insulator classes. However, in

industrial applications, the metric of most interest to application personnel is the recall rate because the level of recall is directly related to whether the model can identify all the insulators in the aerial images.

## 4 Insulator detection based on the improved deep learning algorithm

Traditional location identification and fault diagnosis methods are difficult to adapt to the detection work in the era of big data and operate inefficiently. Because of the ability to extract high-level abstract features from a large amount of data, deep learning techniques are gradually being widely used in the field of location identification and fault diagnosis. In deep learning-based insulator detection tasks, algorithms applied in industrial scenarios must achieve high accuracy and real-time performance. In the inspection process, aerial insulator images are often encountered with poor detection precision due to complex backgrounds and excessive occlusion. In addition, problems such as multiscale objects in a single insulator image and difficulties in detecting small objects lead to missed detection. As a result, researchers in related fields have improved various deep learning algorithms to alleviate the problems encountered in insulator detection. It is important to note that accurate insulator location identification is a prerequisite for insulator fault diagnosis. The insulator location identification is mainly an operation to obtain the location of insulators and their categories in response to the problems of partial obscuration of insulators in inspection images, the difficulty of detection, and low accuracy. On the contrary, the fault diagnosis of insulators is mainly an operation performed to obtain the fault points of insulators caused by the influence of the self-weight of conductors, lightning, wind, ice, snow, dust pollution, and other factors that cause insulators to fall off the string, break, foul and so on. In the improved algorithms mentioned in this section, for insulator location identification and fault diagnosis work, researchers use methods such as one-stage algorithms and two-stage algorithms for deep convolutional neural networks. These algorithms contain both improved algorithms to improve insulator localization recognition rate and improved algorithms to enhance insulator fault diagnosis.

### 4.1 Improving precision and maintaining real-time

A high-voltage insulator requires continuous monitoring and inspection to prevent failures and emergencies (DianaSadykova, 2019). Manual inspection is costly because it requires covering a large geographical area where severe weather conditions can affect the proper operation of the insulators. Automatic detection

of insulators from aerial images is the first step towards performing real-time classifications of insulator conditions using a UAV. The accurate identification of insulators or insulator faults in real time is a difficult problem, and researchers have proposed many improved algorithms from this perspective.

#### 4.1.1 On the improvements of the convolutional network layer

Zhang et al. (2021) increased the convolutional layer perceptual field by replacing the original convolutional layer in the YOLOv3 backbone network with a cavity convolutional layer with an expansion rate of two, which enabled the convolutional network to fuse more object information while ensuring the resolution. According to the morphological characteristics of insulators, the distance metric formula in the k-means clustering algorithm is improved to cluster an anchor box size that is more suitable for insulator characteristics. There is 7.9% improvement in the recognition AP value while operating in real-time. Pan et al. (2020) analysed the scattering transform principle and convolutional neural network (CNN) to make a low-pass filter for scattering coefficient processing. The Gram matrix method is combined to reduce the noise interference of insulator string background information to enhance the edge texture features of low-frequency coefficients. Finally, the SSD network framework achieves the efficiency of CNN for real-time insulator string localization calculations. Ding (2020) proposed three lightweight fully convolutional blocks for use with stacked neural networks on the network structure underlying the YOLO-LITE algorithm. They used the k-means algorithm to cluster the bounding boxes in the training set and proposed a multifeature fusion method and a reduced model input size method. Finally, the best real-time object recognition and localization model comes from various stacking experiments. By improving the convolutional layer, researchers can improve the feature extraction ability of the whole network. Nevertheless, as the convolutional module increases, the detection time consumption increases, and the detection speed decreases accordingly.

#### 4.1.2 On the improvements of the backbone network and feature extraction

Liu et al. (2021c) developed a new deep learning-based intelligent diagnosis method for electrical insulators, termed Box-Point Detector, which consists of a deep convolutional neural network followed by two parallel branches of convolutional heads. The proposed Box-Point Detector forms an efficient end-to-end architecture which implements all predictions including regions and endpoints into a single network and adopts a smaller downsampling rate to generate high-resolution feature maps in order to preserve more original information faults for small sizes. Experimental results show that Box-Point Detector can accurately diagnose high-voltage

insulator faults in real time under a variety of conditions. Chen et al. (2021) used VoVNet with a stronger feature extraction capability as the backbone network of the YOLOv3 algorithm. A novel feature enhancement module was also proposed to effectively enhance shallow feature maps' semantic information and perceptual field—an effective improvement of insulator detection precision of the model with a guaranteed detection speed. Chen and Min (2020) added a new small-target-friendly 4-fold downsampled residual block to the middle of the second residual block and the third residual block of Darknet-53, the backbone network of YOLOv3, to improve the detection precision of small objects. Based on the feature that the positions of insulators in similar images are roughly the same, the images are classified by the perceptual hashing algorithm, and the candidate region scanning strategy is used to speed up the detection speed for similar images. The accuracy of the improved insulator detection method improved from 93.6% to 99.2%, and the detection speed of similar images improved by 4.6%. Wu et al. (2019) used Crop-MobileNet as the base network to extract depth features. They used the Euclidean distance-based K-means algorithm to improve the stability of the generated last frame. Finally, the YOLOv3 architecture is used to detect the insulators and the locations of faults. The method significantly improves the detection speed of the network while maintaining the insulator detection performance and can meet the real-time detection requirements of UAV power inspection. Liu et al. (2021d) relied on the convolutional neural network (CNN) to extract features and process them to form an end-to-end detection network. MobileNet was used as the base network of SSD to achieve high-speed and high-precision detection of insulators. The detection precision of porcelain insulators on 500 kV lines and composite insulators on 220 kV lines reaches 96.29% and 90.85%, respectively, and the average detection speed reaches 43 F·s<sup>-1</sup>, which can meet the requirements of real-time insulator detection. Ji et al. (2020) used the MobileNetV3 network as the feature extraction backbone. The architecture uses the MobileNetV3 network as the feature extraction backbone, designs a new perceptual field module RFB-X, and uses the focal-loss function to solve the positive and negative sample imbalance problem. The experimental results show that the model improves the speed and accuracy of insulator detection. Huang et al. (2022) designed a deep learning network incorporating multidimensional feature extraction to alleviate the problem of poor real-time insulator detection and insufficient feature extraction capabilities based on edge computing. ResNet101 is used as the backbone feature extraction network. The inception module is used to build a data pooling layer and embeds a compressed excitation module and a convolutional attention module to efficiently extract features from different dimensions. The edge recognition framework of insulator states is also constructed using the joint technical means of the cloud-edge and edge-edge federation collaboration. The experiments

demonstrate the effectiveness of the insulator state edge recognition method that fuses multidimensional features. From the perspective of backbone network optimization, researchers fuse lightweight backbone networks to improve the detection speed. The feature extraction capability is improved to maximize the detection precision while maintaining the detection speed.

#### 4.1.3 On the improvements of the activation function of the network layers and the loss function of training

Yao and Qin (2020) used the “GIoU loss function” instead of the loss function of the original YOLOv3 to improve the detection precision of the insulators without increasing the model size and verified the effectiveness of the improved algorithm. Yan and Chen (2020) used the focal loss function and equilibrium cross-entropy function instead of the loss function of the original algorithm. They trained the network analytically, selecting the frozen layer and adopting a multistage migration learning strategy. The study’s results proved that the proposed method has high precision and real-time performance. Ji (2020) modified the bounding box dimension in the network. They improved the original loss function using the Focal Loss function, the linear activation function, and the Leaky ReLU activation function in the network using the Mish function. The improved YOLOv3 calculates a 3.4% improvement in its detection accuracy while maintaining detection speed. The researchers improved the activation and loss functions of the original algorithm according to the actual, especially the Focal Loss function, by reducing the weights of the easily classified samples, making the model focus more on the difficult classified samples during training, achieving an increase in convergence speed as well as more accurate localization detection.

In the literature (Wu et al., 2019; Chen and Min, 2020; Ding, 2020; Ji, 2020; Ji et al., 2020; Pan et al., 2020; Yan and Chen, 2020; Yao and Qin, 2020; Chen et al., 2021; Liu et al., 2021c; Liu et al., 2021d; Zhang et al., 2021; Huang et al., 2022), by improving the selection algorithm, the detection AP or mAP is improved compared to the original algorithm while taking into account the detection speed, which can meet the requirements of real-time detection. However, most of the literature only focuses on improving detection precision while ignoring the problem of missed insulator detection, i.e., the high or low recall rate. Thus, subsequent research can further improve the algorithm to improve the detection recall of the model while maintaining the current results.

## 4.2 Complex background and target occlusion

With the fault of power line insulators leading to transmission system failures, insulator detection systems based

on overhead platforms have been widely used. Insulator targets or diagnostics are performed in the complex context of aerial images, which is an exciting but challenging problem (Tao et al., 2020). Researchers have proposed many improved algorithms from this perspective.

#### 4.2.1 On the improvements of the backbone network and feature extraction

Zhang and Guo (2021) fine-tuned the proportion of RPN candidate regions on the original model to increase the number of anchors. They replaced the backbone network VGG16 in Faster R-CNN with ResNet-101 and used multiscale training for model training. This method can improve the insulator detection precision and alleviate the missed detection problem due to partial insulator occlusion. Compared with the original model, the detection precision of the improved model is improved by 4.88%. Zhu et al. (2021) replaced the backbone VGG16 in Faster R-CNN with ResNet-50 as the backbone network, improved the feature extraction network framework, and performed multiscale feature fusion. The mAP on the test set reaches approximately 98.5%, which alleviates the insufficient recognition accuracy due to the complex background of the insulator images. Tan (2021) used the insulator localization network (one-level detection backbone) of SSD to extract multilevel features and make predictions. Additionally, they introduced a densely connected convolutional network (DenseNet) to enhance the classification capability of the insulator detection system. The researchers optimized the original algorithm backbone network to improve the network’s ability to extract features. It is possible to extract more relevant features of insulators in complex backgrounds.

#### 4.2.2 On the improvements of feature processing

Yi et al. (2021) trained Faster R-CNN with multiscale features. The proportion of candidate regions generated by sliding windows is adjusted according to the insulator’s characteristics. An adversary generation strategy for detecting difficult samples is introduced to detect insulators of different sizes and accurately obscured transmission lines. Huang et al. (2021) proposed a feature pyramid and multitask learning-based insulator detection method. The feature pyramid is constructed by fusing high- and low-dimensional feature information to avoid the loss of detailed information such as object location and achieves efficient detection of insulators in complex backgrounds. A multitask learning algorithm is also introduced to further enhance the generalization ability of the model and improve the insulator detection precision. Chen et al. (2018) proposed an insulator detection method based on a U-net network. The shallow features and high-dimensional features are fused by a superposition method, where the shallow, high-resolution feature maps are used for pixel localization, and the deep high-dimensional feature maps are used for pixel



classification. It avoids the loss of detailed information such as target location and can effectively detect insulators in complex backgrounds. Liu et al. (2021e) proposed an improved algorithm, YOLOv3-dense, which enhances the reuse and propagation of features. Additionally, a multilayer feature mapping module is used in the YOLOv3-dense network to obtain the rich semantic information in the upper and lower layers, which has good performance for detecting insulators of different sizes under different background disturbances. Wang et al. (2021) synergized the full convolutional network (FCN) with the YOLOv3 object detection algorithm. First, the FCN algorithm is used to achieve the initial segmentation of insulator targets to avoid the interference of background regions on the insulator fault detection. Second, the YOLOv3 model is used for insulator fault detection, drawing on the idea of aiding in labelling and predicting the category of insulator faults on the output tensor of three scales to ensure that the model detects insulator faults of different sizes accurately. The K-means clustering algorithm is used to optimize the anchor box parameters of YOLOv3. Compared with improving the backbone network, the researchers choose to optimize from the perspective of feature fusion by superimposing and fusing multiple layers of features to extract the more insulator features.

#### 4.2.3 On the improvements of anchor generation and filter, the loss function of training, and the activation function of the network layers

Zhao et al. (2019a) improved the anchor generation method and nonmaximal suppression (NMS) screening method by Faster R-CNN. At the same time, the improved method has significantly improved the detection of insulators of different scales and can effectively distinguish and detect insulator shading. Gao and Wang (2022) proposed an improved algorithm based on the CornerNet-Lite network model. The algorithm applies the LeakyReLU function to design a more reasonable activation function. It can effectively alleviate the problem of model leakage when pylons intermittently block insulators, and multiple targets are clustered. Tian et al. (2021) incorporated the squeeze-and-excitation (SE) attention module in the YOLOv5s network to strengthen the network's ability to identify insulator objects. K-means clustering is also used to construct the prior frame of insulators to improve the localization precision. Last, a loss function with joint confidence and localization tasks is built to enhance the performance of insulator detection in complex backgrounds. Tang et al. (2021) combined the object segmentation model SERes-Unet with the improved YOLOv5. The SERes-Unet model cuts the images into two categories, the background and insulator targets. The improved YOLOv5 algorithm is responsible for detecting these two classes and filtering by nonmaximal suppression to obtain the insulator positions in the full complex background. The researchers' optimization

makes the improved regression frame more adaptable to insulator objects with occlusion in real scenes.

#### 4.2.4 On the improvements of the training strategy of the network

Zhao et al. (2019b) improved the regional fully convolutional network (R-FCN). The aspect ratio of the proposed frame of the RPN in the R-FCN model is modified according to the aspect ratio characteristics of the insulator object. Additionally, for the masking problem, an Adversarial Spatial Dropout Network (ASDN) layer is introduced into the R-FCN model to generate masks for some positions of the feature map to obtain incomplete samples of the object features to improve the model's detection performance for samples with poor object features. Miao et al. (2019) proposed a method combining SSD with a two-stage fine-tuning strategy to automatically extract the multilevel features from aerial images. The two-stage fine-tuning strategy is also implemented using different training sets, enabling the well-trained SSD model to directly and accurately identify insulators in aerial images with complex backgrounds. Lai et al. (2019) combined edge computing, line detection, image rotation and vertical projection. The method learns the features of various insulators in complex backgrounds by training the YOLOv2 network to achieve accurate identification. Optimizing the training strategy of the network aims to improve the network's adaptability to extract features in various complex backgrounds and to have more vital generalization abilities in different application scenarios.

Most studies (Chen et al., 2018; Zhao et al., 2019a; Zhao et al., 2019b; Lai et al., 2019; Miao et al., 2019; Liu et al., 2021e; Huang et al., 2021; Tan, 2021; Tang et al., 2021; Tian et al., 2021; Wang et al., 2021; Yi et al., 2021; Zhang and Guo, 2021; Zhu et al., 2021; Gao and Wang, 2022) use AP as the main evaluation index, which is improved compared with the original algorithm. However, the model may not recognize insulator objects fully in application scenarios with complex images and object occlusion backgrounds, leading to partial insulator missed detection. In the literature (Chen et al., 2018; Zhao et al., 2019a; Zhao et al., 2019b; Lai et al., 2019; Miao et al., 2019; Liu et al., 2021e; Huang et al., 2021; Tan, 2021; Tang et al., 2021; Tian et al., 2021; Wang et al., 2021; Yi et al., 2021; Zhang and Guo, 2021; Zhu et al., 2021; Gao and Wang, 2022), only half of them also use the recall rate as the evaluation index, and the vast majority are not mainly aimed at improving the recall rate. Therefore, researchers can work on improving the detection recall based on the application algorithm for this scenario.

### 4.3 Multiscale and small target

Insulator failures may endanger the safety of the entire transmission system. Therefore, monitoring insulators is a priority to maintain the safe operation of power systems.

However, insulator targets or insulator defects in insulator images may have different sizes, and it is still a challenging problem to improve the detection precision of small targets at present (Gao et al., 2021a). Meanwhile, small target detection is a hot research topic, especially in the electric power field. For this problem, researchers have also proposed many improved algorithms.

#### 4.3.1 On the improvements of feature processing and anchor box scaling

Shen et al. (2020) fused the shallow feature map with the deep feature map in Faster R-CNN to improve the algorithm's ability to extract object features. The scale of the anchor frame is improved according to the shape characteristics of insulators to enhance the detection ability of small-scale insulators. Multiscale training is also performed to reduce the influence of insulators of different scales on the recognition rate. Zhou et al. (2020) also used the multiscale feature fusion method. The backbone network VGG16 of Faster R-CNN was also replaced with ResNet-101 to effectively improve small-scale object detection precision. On the other hand, Jiang et al. (2022) combined the feature pyramid network (FPN) with the Faster R-CNN algorithm and fused multiscale features. At the same time, the algorithm improves the maximum pooling layer. It uses the soft-NMS algorithm instead of NMS to circumvent the case of mistakenly deleting overlapping detection frames due to an overlapping target occlusion, which reduces the leakage of small-sized insulators in images. Wu (2020) achieved multilayer network feature fusion using a Concat connection. The improved YOLOv3 algorithm was also trained using an end-to-end multiscale training approach. The improved algorithm has a significant improvement in the detection precision and detection speed for small objects. Chen et al. (2020) proposed a YOLOv3-based feature selection network (FS-YOLOv3) detection method. The redundant low-level detail features are filtered out using a pyramidal feature attention network, and the low-level detail features are fused with deep-level semantic features. It effectively alleviates insulator missed detection and inaccurate localization due to the small proportion of insulators and the complex background in the infrared power images. Pan (2019) proposed an MFIDN insulator detection network model based on multiscale and fine-grained features. The model shows that fine-grained features and multiscale prediction positively enhance the insulator detection by comparing the single model structure and the single output of the one-stage detection algorithm. Yan and Wang (2021) proposed a method to detect insulator rust faults based on balanced feature fusion SSD. For the problem of a missed detection of small objects in insulator rust objects, the feature layers in the backbone network of SSD are fused with balanced features. This feature fusion method improves the network's ability to capture detailed information and solves part of the small target miss detection problem. For multiscale object detection, researchers choose to optimize from the perspective of multiscale feature fusion and multiscale anchor frame reconstruction. The multilayer features are superimposed and fused to achieve more

TABLE 2 Algorithm improvement strategies for the comprehensive literature.

#### No. Improvement strategies

- |      |   |
|------|---|
| (1)  | Modify the convolution layer parameters   |
| (2)  | Improving the maximum pooling layer   |
| (3)  | Using the Gram matrix method to reduce background noise   |
| (4)  | Multiscale feature fusion   |
| (5)  | Adding attention mechanism module   |
| (6)  | Increase the network width  |
| (7)  | Replace the loss function   |
| (8)  | Replace the activation function   |
| (9)  | Optimize Anchor scaling/number/generation method  |
| (10) | Optimize the screening method of Anchor (nonmaximum suppression-NMS)  |
| (11) | Modify backbone network   |
| (12) | Improve the feature extraction network fetching framework   |
| (13) | Adding feature enhancement module   |
| (14) | Classification of similarly located insulators using a perceptual hashing algorithm   |
| (15) | Training the network using a multistage migration learning strategy   |
| (16) | Automatic adjustment of learning rate using cosine annealing learning rate decay method in the training phase                                     |
| (17) | Training the network using multiscale   |
| (18) | Introduction of feature reuse network (DenseNet)  |
| (19) | Introduction of adversary generation strategy to detect difficult samples   |
| (20) | Introduction of adversarial space discarding network to generate incomplete samples of target features (generation of defective samples)          |
| (21) | Introduction of multitask learning algorithm  |
| (22) | Fusion of the two algorithms  |
| (23) | The combination of the target segmentation algorithm and target detection algorithm   |
| (24) | One-stage algorithm combined with a two-stage algorithm fine-tuning strategy  |
| (25) | Optimization of the training strategy of the network by combining edge computing, line detection, image rotation, and vertical projection methods |
| (26) | Test phase to generate high-quality images for the presence of small targets using hypersegmentation generation network                           |

insulator detail extraction. Meanwhile, multiscale anchor frame reconstruction can obtain more suitable regression frames. However, it also increases the computational time consumption of the network, thus reducing the detection speed.

#### 4.3.2 On the improvements of the network's training strategy and convolutional layer combinations

Gao et al. (2021b) proposed an improved insulator fault diagnosis method for transmission lines with YOLOv4. A multistage migration learning strategy and cosine annealing learning rate decay method are used in the training phase to improve the training speed and overall performance of the network. Meanwhile, in the testing phase, the images with small objects are tested after generating high-quality images

TABLE 3 Evaluation index of detection performance of comprehensive literature.

Literature	Improved algorithm	Application scenario	Location identification	Fault diagnosis	AP <sup>O</sup> / %	AP <sup>I</sup> / %	mAP <sup>O</sup> / %	mAP <sup>I</sup> / %	R <sup>O</sup> / %	R <sup>I</sup> / %	FPS <sup>O</sup> / (f-s-1)	FPS <sup>I</sup> / (f-s-1)	Improvement strategies
Zhang et al. (2021)	YOLOv3	(a)	√		83.90	91.80	—	—	80.30	89.10	33.00	35.00	(1) (9)
Pan et al. (2020)	SSD	(a)	√		—	—	—	—	—	89.56	—	31.00	(3)
Ding, (2020)	YOLO-LITE	(a)	√		32.98	40.65	—	—	48.00	52.00	15.00	25.00	(4) (9)
Liu et al. (2021c)	Box-Point Detector (New)	(a)		√	71.90	73.70	—	—	—	—	—	30.30	(11) (12)
Chen et al. (2021)	YOLOv3	(a)	√		—	—	91.94	98.01	82.74	97.67	34.00	29.00	(11) (13)
Chen and Min, (2020)	YOLOv3	(a)	√		93.60	99.20	—	—	—	—	—	—	(13) (14)
Wu et al. (2019)	YOLOv3	(a)		√	—	—	90.00	84.00	—	—	26.10	40.90	(9) (11)
Liu et al. (2021d)	SSD	(a)	√		—	(procelain: 96.29) (composite: 90.85)	—	93.57	—	(procelain: 95.12) (composite: 85.24)	—	35~43	(11)
Ji et al. (2020)	RFB	(a)	√		—	—	91.36	91.91	—	—	11.20	12.90	(7) (11)
Huang et al. (2022)	ResNet101	(a)	√		97.50	99.70	—	—	97.50	98.90	—	18.6~23.6	(4) (5)
Yao and Qin, (2020)	YOLOv3	(a)	√		90.00	92.50	—	—	—	—	58.80	58.80	(7)
Yan and Chen, (2020)	YOLOv3	(a)	√		89.20	91.80	—	—	89.60	90.10	—	—	(7) (15)
Ji, (2020)	YOLOv3	(a)		√	—	—	89.40	93.10	—	—	25.00	24.00	(7) (8) (9)
Zhang and Guo, (2021)	Faster R-CNN	(b)	√		86.82	91.70	—	—	74.67	84.67	—	—	(9) (11) (17)
Zhu et al. (2021)	Faster R-CNN	(b)	√		—	—	—	98.50	—	—	—	—	(4) (11) (12)
Tan, (2021)	SSD	(b)	√		—	95.00	—	—	—	92.00	—	10.00	(18)
Yi et al. (2021)	Faster R-CNN	(b)	√		89.11	93.44	—	—	77.75	85.50	—	—	(9) (17) (19)
Huang et al. (2021)	FPN	(b)	√		80.10	92.30	—	—	80.50	90.10	—	—	(4) (21)
Chen et al. (2018)	U-net	(b)	√		85.80	88.90	—	—	—	—	—	—	(4)
Liu et al. (2021e)	YOLOv3	(b)	√		90.31	94.47	—	—	91.00	96.00	—	—	(4) (18)
Wang et al. (2021)	FCN + YOLOv3	(b)		√	—	—	92.23	96.88	—	—	—	—	(9) (23)
Zhao et al. (2019a)	Faster R-CNN	(b)	√		64.00	81.80	—	—	—	—	—	—	(9) (10)
Gao and Wang, (2022)	CornerNet-Lite	(b)	√		87.00	93.00	—	—	—	—	—	—	(8)
Tian et al. (2021)	YOLOv5	(b)	√		—	—	98.24	98.46	—	—	—	—	(5) (7) (9)
Tang et al. (2021)	U-net, YOLOv5	(b)		√	95.30	97.30	—	—	99.00	99.00	—	—	(10) (23)
Zhao et al. (2019b)	R-FCN	(b)	√		77.27	84.29	—	—	—	—	—	—	(9) (20)

(Continued on following page)

TABLE 3 (Continued) Evaluation index of detection performance of comprehensive literature.

Literature	Improved algorithm	Application scenario	Location identification	Fault diagnosis	AP <sup>O</sup> / %	AP <sup>I</sup> / %	mAP <sup>O</sup> / %	mAP <sup>I</sup> / %	R <sup>O</sup> / %	R <sup>I</sup> / %	FPS <sup>O</sup> / (f-s-1)	FPS <sup>I</sup> / (f-s-1)	Improvement strategies
Miao et al. (2019)	SSD	(b)	√		81.28	94.12	72.77	90.41	79.00	90.00	—	—	(24)
Lai et al. (2019)	YOLOv2	(b)		√	—	90.00	—	—	—	—	—	30.00	(25)
Shen et al. (2020)	Faster R-CNN	(c)	√		87.05	95.80	86.80	93.60	85.25	92.35	—	4.20	(4) (9) (17)
Zhou et al. (2020)	Faster R-CNN	(c)		√	—	—	87.30	91.30	—	—	—	—	(4) (11)
Jiang et al. (2022)	Faster R-CNN	(c)	√		—	—	50.63	81.97	—	—	—	—	(2) (4) (10) (22)
Wu, (2020)	YOLOv3	(c)		√	93.15	95.31	92.51	94.72	—	—	45.93	46.12	(4) (17)
Chen et al. (2020)	YOLOv3	(c)	√		—	—	—	—	93.21	95.54	17.2	16.8	(4) (5)
Pan, (2019)	YOLO + SSD	(c)		√	83.10	93.20	—	—	—	—	—	—	(4)
Yan and Wang, (2021)	SSD	(c)	√		78.20	80.20	—	—	—	—	—	—	(4)
Gao et al. (2021b)	YOLOv4+SR-GAN	(c)		√	—	—	—	89.54	—	—	—	12.24	(15) (16) (26)
Yang, (2019)	Faster R-CNN	(c)	√		86.50	88.90	—	—	—	—	—	—	(4) (5)
Zuo et al. (2019)	Adaboost	(c)	√		85.00	95.00	—	—	86.00	93.00	—	—	(1) (4)

using the hyperscore generation network to improve the ability to identify small objects effectively. Yang (2019) added the conv3 layer feature map on top of the single-scale feature map and embedded the STN (Spatial Transformer Networks) module in the conv3 layer feature map. The improved algorithm can detect the image's small object insulators and rotating insulators better. Zuo et al. (2019) proposed a cross-connected convolutional neural network-based insulator detection method. The method connects the last three convolutional layers of the region suggestion network to the fully connected layer separately. This connection allows the convolutional features of these three layers to be fed into the classifier and regression layers simultaneously, resulting in a series of high-quality insulator candidate regions. Finally, the obtained region of interest features are fed into the cascaded Adaboost classifier, which can effectively identify and precisely locate insulators at different scales. The researchers worked on changing the training strategy of convolutional networks and the way convolutional layers are combined. This improvement reduces the traditional multiscale feature fusion time to a certain extent while achieving the purpose of multiscale feature extraction.

Most of the literature (Pan, 2019; Yang, 2019; Zuo et al., 2019; Chen et al., 2020; Shen et al., 2020; Wu, 2020; Zhou et al., 2020; Gao et al., 2021b; Yan and Wang, 2021; Jiang et al., 2022) uses AP or mAP as the primary evaluation metric, which is improved compared to the original algorithm. However, focusing on the recall rate in the application scenario of multiscale object and small object detection is more important. Only a few papers in the literature (Pan, 2019; Yang, 2019; Zuo et al., 2019; Chen et al., 2020; Shen et al., 2020; Wu, 2020; Zhou et al., 2020; Gao et al., 2021b; Yan and Wang, 2021; Jiang et al., 2022) also use the recall rate as an evaluation metric. Multiscale insulator object detection is prone to the missed detection of some insulator objects of different scales. Therefore, the study of algorithm improvement for this application scenario should focus on improving the recall rate and thus reducing the missed insulator targets of different scales. Hence, the recall rate can be used as the main evaluation index in subsequent research for this scenario.

## 5 Comprehensive literature experimental results and conclusion

### 5.1 Experimental results

Literature review shows that most of the research performed in the past mainly addresses three application situations: Scenario (a): improved detection precision and real-time detection, Scenario (b): complex background and target occlusion of images, and Scenario (c): multiscale and small object detection difficulties. For the three common application scenarios, the researchers made targeted improvements from the perspective of the original deep learning

algorithm. The results produced were all improved compared with the original algorithm. A comparison of the algorithm performance evaluation indices is shown in Table 3.  $AP^O$ ,  $mAP^O$ ,  $R^O$ , and  $FPS^O$  represent the performance indicators of the original algorithm.  $AP^I$ ,  $mAP^I$ ,  $R^I$ , and  $FPS^I$  represent the performance indicators of the improved algorithm. Meanwhile, the algorithm improvement strategies of the related literature are organized as shown in Table 2. They are arranged by number. The algorithm improvement strategies of the corresponding literature can also be seen in the "Improvement strategies" in Table 3.

### 5.2 Conclusion

By generalizing the experimental data from the literature and combining them with different application scenarios, there are following conclusions can be drawn:

For application Scenario (a), from the data in Table 3 literature (Wu et al., 2019; Chen and Min, 2020; Ding, 2020; Ji, 2020; Ji et al., 2020; Pan et al., 2020; Yan and Chen, 2020; Yao and Qin, 2020; Chen et al., 2021; Liu et al., 2021c; Liu et al., 2021d; Zhang et al., 2021; Huang et al., 2022), most researchers choose the one-stage YOLO family of algorithms for improvement, which is closely related to its advantages. The YOLO algorithm is fast and can perform real-time detection because the YOLO framework treats target detection as a regression problem, unlike the RCNN algorithm's complex processing process. However, the YOLO algorithm can be said to trade precision for speed, and the localization precision of detected objects is poor. Therefore, many researchers choose the YOLO algorithm to improve the detection speed while ensuring the detection speed to achieve the double improvement of detection precision and speed. The SSD algorithm combines R-CNN's anchor mechanism and YOLO's regression idea, which can also meet the scenario requirements through improvement. Thus, the YOLO series and SSD algorithm are preferred to improve the detection precision and speed.

For application scenarios (b) and (c), the data from the literature (Chen et al., 2018; Zhao et al., 2019a; Zhao et al., 2019b; Lai et al., 2019; Miao et al., 2019; Pan, 2019; Yang, 2019; Zuo et al., 2019; Chen et al., 2020; Shen et al., 2020; Wu, 2020; Zhou et al., 2020; Gao et al., 2021b; Liu et al., 2021e; Huang et al., 2021; Tan, 2021; Tang et al., 2021; Tian et al., 2021; Wang et al., 2021; Yan and Wang, 2021; Yi et al., 2021; Zhang and Guo, 2021; Zhu et al., 2021; Gao and Wang, 2022; Jiang et al., 2022) in Table 3 show that the algorithms chosen by the researchers are not as high as those required to meet the real-time requirements. Among the selected algorithms are mostly the Faster R-CNN algorithm for two-stage detection, which is related to its advantages. Faster RCNN contains RPN networks that enable a high accuracy detection performance. At the same time, compared with other one-stage detection networks, the two-stage network is more precise and can solve multiscale and small-object problems

more effectively. The corresponding YOLO algorithm is also selected for improvement according to the researchers' implementation needs (detection speed, etc.).

As seen from the data in Table 3, researchers performed algorithm optimization for three application scenarios, mainly with the primary goal of improving the detection accuracy (AP/mAP). In practical applications, focusing on accuracy and ignoring the recall rate is easy, while a low recall rate may directly lead to target misdetection. Compared with low object recognition precision, the loss caused by the low recall rate of missed detection may be greater. Therefore, the detection recall rate should be the primary evaluation index in the insulator detection task of intelligent power grids.

## 6 Summary and prospects

This paper focuses on a review of insulator detection techniques for intelligent power grids based on deep learning. First, the development of deep learning-based object detection algorithms is summarized. Second, we review the implementation of deep learning algorithms on insulator detection tasks and improved algorithms for three application scenarios. As seen from Table 3, for application scenarios (a), (b), and (c), researchers have improved the existing object detection networks based on them, and all have achieved some success in terms of experimental results.

Overall, the detection speed of the regression-based object detection algorithm on the insulator detection task is relatively fast, which can meet the requirements of real-time detection, but the detection precision is relatively low. Therefore, improving detection precision based on maintaining the detection speed is a hot research topic. The candidate region-based object detection algorithm outperforms the regression-based object detection algorithm's detection precision performance. The improved algorithm for scenarios (b) and (c) is effective, but the detection speed has difficulty meeting the real-time requirements and faces challenges in application promotion.

By summarizing and analysing the above technologies, we list the current challenges encountered in insulator location identification or fault diagnosis and possible future research trends, while expecting to promote research development in this field.

(1) Firstly, most of the current insulator fault diagnosis techniques based on deep learning are based on locating and identifying insulator images and then diagnosing the faults in insulators. This method is complicated and time-consuming in practical application. At the same time, the fault diagnosis effect of insulator is limited by the localization and recognition effect of insulator. Secondly, there are many types of faults in insulators, such as "cracks and corrosion." These defects have obscure features and are difficult to train directly as a dataset. This is because the models trained using these datasets will misdiagnose

similar features of non-insulator defects. Therefore, one of the next research trends is how to directly diagnose potential faults in insulators and determine whether the faults belong to insulators or not. For example, if insulators and insulator defect locations are used as separate data sets for training the model. Then, in the localization and identification stage of insulator target and defect location, the defect location detected by the model and the insulator location detected are calculated by area overlap, and the location of the fault point of the insulator can be initially diagnosed.

(2) There is insufficient sample data for insulator faults. Also, there are many types of faults in insulators and the sample data of the fault types that can be collected are unbalanced. And in model training, insufficient data can lead to overfitting of the model. Unbalanced data types can cause false training accuracy of the model, which can lead to false detection (Cai et al., 2022). Therefore, without being able to change this hard condition, the training strategy of the model should be changed, and the migration learning technique will be one of the key methods to solve the problem of insufficient insulator fault samples.

## Author contributions

DC wrote the first draft of the manuscript and performed the experiments. ZyZ reviewed the manuscript and provided the protocol. ZZ assisted in the analysis. All the authors contributed to manuscript's revision, and read and approved the submitted version.

## Funding

The research was supported by the project of Dongguan Sci-tech Commissioner Program (20221800500082).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). "SegNet: A deep convolutional encoder-decoder architecture for image segmentation[J]." in IEEE Transactions on Pattern Analysis & Machine Intelligence, 01 December 2017 (IEEE), 1.
- Bochkovskiy, A., Wang, C. Y., and Liao, H. (2020). YOLOv4: Optimal speed and accuracy of object detection. arXiv.
- Cai, D., Zhang, Z., and He, G. (2022). "A study on the effectiveness of detection of unbalanced datasets based on faster R-CNN," in A Study on the Effectiveness of Detection of Unbalanced Datasets Based on Faster R-CNN, 2022 12th International Conference on Power, Shiga, Japan, 25-27 February 2022 (IEEE).
- Cai, G., Wang, Y., and Zhou, M. (2018). Unsupervised domain adaptation with adversarial residual transform networks. arXiv.
- Chen, J., Zhou, X., and Zhang, R. (2018). Aerial insulator detection based on U-net network[J]. *J. Shaanxi Univ. Sci. Technol.* 36 (4), 5. doi:10.3969/j.issn.1000-5811.2018.04.028
- Chen, K., Shi, L., and Liu, B. (2021). Improved YOLOv3 method for transmission line insulator detection[J]. *Sci. Technol. Innovation Appl.* 11 (34), 5. doi:10.19981/j.CN23-1581/G3.2021.34.018
- Chen, M., Zhao, L., and Yuan, L. (2020). An infrared image insulator detection method based on feature selection YOLOv3 network [J]. *Infrared Laser Eng.* 2, 6. doi:10.3788/IRLA20200401
- Chen, Z., and Min, F. (2020). Contact network insulator detection method based on improved YOLO V3[J]. *J. Wuhan Univ. Eng.* 42 (4), 5. doi:10.19843/j.cnki.CN42-1779/TQ.201906027
- Csurka, G. (2017). Domain adaptation for visual applications: A comprehensive survey. arXiv.
- Dai, J., Li, Y., and He, K. (2016). R-FCN: Object detection via region-based fully convolutional networks. arXiv.
- Diana, S., Sadykova, D., Bagheri, M., and James, A. (2019). IN-YOLO: Real-Time detection of outdoor high voltage insulators using UAV imaging[J]. *IEEE Trans. Power Deliv.* 35 (3), 1599–1601. doi:10.1109/TPWRD.2019.2944741
- Ding, G. (2020). *Research on real-time insulator identification and localization method in aerial images[D]*. Beijing: North China Electric Power University.
- Dosovitskiy, A., Beyer, L., and Kolesnikov, A. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv.
- Eigen, D., and Fergus, R. (2014). "Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture," in 2015 IEEE International Conference on Computer Vision (ICCV), CentroParque Convention Center in Santiago, Chile, December 13–16, 2015 (IEEE).
- Ganin, Y., and Lempitsky, V. (2014). Unsupervised Domain Adaptation by Backpropagation. arXiv.
- Gao, Q., and Wang, M. (2022). Deep learning algorithm based aerial insulator detection. *Electrotechnology.* 21, 1033. doi:10.3390/s21041033
- Gao, W., Zhou, C., and Guo, M. (2021). Research on insulator defect identification based on improved YOLOv4 and SR-GAN [J]. *J. Electr. Mach. Control* 25 (11), 12. doi:10.15938/j.emc.2021.11.011
- Gao, Z., Yang, G., and Li, E. (2021). Novel feature fusion module based detector for small insulator defect detection[J]. *IEEE Sensors J.* 99, 1. doi:10.1109/JSEN.2021.3073422
- Ge, Z., Liu, S., and Wang, F. (2021). Yolox: Exceeding YOLO series in 2021. arXiv.
- Girshick, R., Donahue, J., and Darrell, T. (2013). "Rich feature hierarchies for accurate object detection and semantic segmentation[J]," in Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation, Columbus, OH, USA, 23-28 June 2014 (IEEE).
- Girshick, R. (2015). Fast R-CNN. arXiv.
- He, K., Zhang, X., and Ren, S. (2016). Deep residual learning for image recognition. arXiv.
- He, K., Zhang, X., and Ren, S. (2015). "Spatial pyramid pooling in deep convolutional networks for visual recognition[J]," in IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE).
- Huang, D., Wang, Y., and Hu, A. (2022). A method for insulator state edge identification by fusing multidimensional features[J]. *China Electr. Power* 55 (1), 9. doi:10.11930/j.issn.1004-9649.202011120
- Huang, J., and Zhang, G. (2020). A review of target detection algorithms for deep convolutional neural networks[J]. *Comput. Eng. Appl.* 56 (17), 12. doi:10.3778/j.issn.1002-8331.2005-0021
- Huang, L., Zhao, K., Li, J., and Feng, H. (2021). Insulator image detection based on feature pyramid and multi-task learning[J]. *Electr. Meas. Instrum.* 58 (04), 37–43. doi:10.19753/j.issn1001-1390.2021.04.006
- Hudson, D. A., and Zitnick, C. L. (2021). Generative adversarial transformers. arXiv.
- Ji, Q. (2020). *Research on aerial insulator missing detection based on YOLOv3[D]*. Shenyang: Shenyang Agricultural University.
- Ji, Z., Zhang, G., and Lu, Q. (2020). Real-time insulator identification and localization method based on sensory field module[J]. *Electrotech. Electr.* 9, 5.
- Jia, D., Wei, D., Socher, R., Li, L. J., Fei, F., Li, K., et al. (2009). "ImageNet: A large-scale hierarchical image database," in 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20-25 June 2009 (IEEE), 248–255.
- Jiang, S., Sun, Y., and Yan, D. (2022). Insulator identification based on deep learning algorithm for aerial inspection images[J]. *J. Fuzhou Univ. Nat. Sci. Ed.*, 11, 4647. doi:10.3390/app11104647
- Kampffmeyer, M., Salberg, A. B., and Jenssen, R. (2016). "Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks[C]," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Las Vegas, NV, USA, 26 June 2016 - 01 July 2016 (IEEE).
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks[J]. *Commun. ACM.* 60 (6), 84–90. doi:10.1145/3065386
- Lai, Q., Yang, J., and Tan, B. (2019). Automatic insulator identification and defect diagnosis model based on YOLOv2 network[J]. *China Electr. Power* 52 (7), 9. doi:10.11930/j.issn.1004-9649.201806102
- Li, Y., Zhang, K., and Cao, J. (2021). LocalViT: Bringing locality to vision transformers. arXiv.
- Li, Z., Peng, C., and Yu, G. (2017). Light-head R-CNN: In defense of two-stage object detector. arXiv.
- Lin, T. Y., Goyal, P., and Girshick, R. (2017). "Focal loss for dense object detection [J]," in IEEE Transactions on Pattern Analysis & Machine Intelligence, Venice, Italy, October 22–29, 2017 (IEEE), 2999–3007.
- Liu, C., Wu, Y., Liu, J., and Sun, Z. (2021). Improved YOLOv3 network for insulator detection in aerial images with diverse background interference. *Electronics* 10 (7), 771. doi:10.3390/electronics10070771
- Liu, W., Anguelov, D., and Erhan, D. (2016). *Ssd: Single Shot MultiBox detector*. Cham: Springer.
- Liu, X., Miao, X., and Jiang, H. (2021). "Box-point detector: A diagnosis method for insulator faults in power lines using aerial images and convolutional neural networks[J]," in IEEE Transactions on Power Delivery (IEEE), 1.
- Liu, X., Miao, X., and Zhuang, S. (2021). Insulator detection based on lightweight deep convolutional neural network[J]. *J. Fuzhou Univ. Nat. Sci. Ed.* 7, 187–197. doi:10.7631/j.issn.1000-2243.20345
- Liu, Y., Sangineto, E., and Bi, W. (2021). Efficient training of visual transformers with small-size datasets. arXiv.
- Liu, Z., Lin, Y., and Cao, Y. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. arXiv.
- Long, M., and Wang, J. (2015). Learning transferable features with deep adaptation networks. arXiv.
- Ma, L. (2021). *Research on insulator detection method based on feature learning [D]*. Beijing: Beijing University of Technology.
- Miao, X., Liu, X., and Chen, J. (2019). "Insulator detection in aerial images for transmission line inspection using single Shot multibox detector [J]," in IEEE Access (IEEE), 1.
- Nguyen, V. N., Jenssen, R., and Roverso, D. (2018). 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.* 99 (JUL), 107–120. doi:10.1016/j.jepes.2017.12.016
- Pan, C., Shen, P. F., and Zhang, Z. (2020). Research on real-time insulator string positioning based on UAV inspection images[J]. *Electr. porcelain Light. arrester* 1, 7. doi:10.16188/j.isa.1003-8337.2020.01.039
- Pan, Z. (2019). *Research on aerial inspection image insulator detection and fault identification based on deep learning[D]*. Taiyuan: Taiyuan University of Technology.
- Redmon, J., Divvala, S., and Girshick, R. (2016). You only Look once: Unified, real-time object detection. arXiv.

- Redmon, J., and Farhadi, A. (2017). "YOLO9000: Better faster, stronger," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017 (IEEE), 6517–6525. doi:10.1109/cvpr.2017.690
- Redmon, J., and Farhadi, A. (2018). YOLOv3: An incremental improvement. arXiv.
- Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (6), 1137–1149. doi:10.1109/tpami.2016.2577031
- Sener, O., Song, H., and Saxena, A. (2016). Learning transferrable representations for unsupervised domain adaptation. arXiv.
- Shen, Z., Niu, P., and Shi, H. (2020). Fusing multilayer convolutional features for insulator target detection[J]. *China Sci. Technol. Pap.* 15 (7), 7. doi:10.3969/j.issn.2095-2783.2020.07.031
- Tan, J. (2021). Automatic insulator detection for power line using aerial images powered by convolutional neural networks. *J. Phys. Conf. Ser.* 1748 (411), 042012. doi:10.1088/1742-6596/1748/4/042012
- Tang, S., Xiong, H., and Huang, R. (2021). Insulator mask acquisition and defect detection based on improved U-Net and YOLOv5[J]. *Data Acquis. Process.* 36 (5), 9. doi:10.16337/j.1004-9037.2021.05.019
- Tao, X., Zhang, D., Wang, Z., Liu, X., Zhang, H., and Xu, D. (2020). Detection of power line insulator defects using aerial images analyzed with convolutional neural networks. *IEEE Trans. Syst. Man. Cybern. Syst.* 50 (4), 1486–1498. doi:10.1109/tsmc.2018.2871750
- Tian, Q., Hu, R., and Li, Z. (2021). Insulator detection based on SE-YOLOv5s[J]. *J. Intelligent Sci. Technol.* 3 (3), 10. doi:10.11959/j.issn.2096-6652.202132
- Tong, L. (2022). *Research on autonomous patrol and insulator identification and positioning method of UAV [D]*. Fuxin: Liaoning Technical University.
- Wang, D. (2021). "Research on insulator defect detection method based on convolutional neural network," in Research on insulator defect detection method based on convolutional neural network, Dalian, China, 29–31 October 2021 (IEEE).
- Wang, Z., Wang, Y., and Wang, Q. (2021). A two-stage insulator fault detection method based on collaborative deep learning[J]. *J. Electr. Eng. Technol.* 36, 3594–3604. doi:10.19595/j.cnki.1000-6753.tces.201320
- Wu, H., Xiao, B., and Codella, N. (2021). CvT: Introducing convolutions to vision transformers. arXiv.
- Wu, R. (2020). *Research on insulator target detection and fault identification based on deep learning[D]*. Anshan: Anhui University of Technology.
- Wu, T., Wang, W., and Yu, L. (2019). Lightweight YOLOV3 method for insulator defect detection[J]. *Comput. Eng.* 45 (8), 6. doi:10.19678/j.issn.1000-3428.0053695
- Xu, W., Xu, Y., and Chang, T. (2021). Co-scale conv-attentional image transform. arXiv.
- Yan, G., and Wang, T. (2021). Insulator detection method based on balanced feature fusion SSD [J]. *China Sci. Technol. Inf.* 10, 2. doi:10.3969/j.issn.1001-8972.2021.10.033
- Yan, H., and Chen, J. (2020). An improved YOLOv3-based insulator string location and state identification method[J]. *High. Volt. Technol.* 10 (2), 771. doi:10.3390/electronics10070771
- Yang, Y. (2019). "Research on target detection algorithm based on convolutional neural network[D]," in 2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA), Shenyang, China, 22–24 January 2021 (IEEE).
- Yao, L., and Qin, Y. (2020). *Insulator Detection Dased on GIOU-YOLOv3[C]//2020 Chinese Automation Congress (CAC)*. IEEE, 5066–5071.
- Yi, J., Chen, C., and Gong, G. (2021). Aerial insulator detection for transmission line based on improved Faster RCNN [J]. *Comput. Eng.* 47 (6), 8. doi:10.19678/j.issn.1000-3428.0059872
- Yuan, K., Guo, S., and Liu, Z. (2021). Incorporating convolution designs into visual transformers. arXiv.
- Yuan, L., Chen, Y. C., and Wang, T. (2021). Tokens-to-Token ViT: Training vision transformers from scratch on ImageNet. arXiv.
- Zhang, H., Li, S., Zhou, H., Fan, X., Lin, Z., Xue, H., et al. (2021). Quick evaluation of lower leg ischemia in patients with peripheral arterial disease by time maximum intensity projection CT angiography: A pilot study. *BMC Med. Imaging* 35 (10), 7. doi:10.1186/s12880-020-00537-5
- Zhang, T., and Guo, Z. (2021). Self-powered all-inorganic perovskite photodetectors with fast response speed. *Nanoscale Res. Lett.* 28 (10), 6. doi:10.1186/s11671-020-03460-4
- Zhao, Z., Cui, Y., and Qi, Y. (2019). Improved insulator detection method based on R-FCN aerial inspection images in line inspection. *Comput. Sci.* 46 (3), 159–163. doi:10.11896/j.issn.1002-137X.2019.03.024
- Zhao, Z., Zhen, Z., Zhang, L., Qi, Y., Kong, Y., and Zhang, K. (2019). Insulator detection method in inspection image based on improved faster R-CNN. *Energies* 12 (7), 1204. doi:10.3390/en12071204
- Zhou, Z., Zhao, C., and Fan, P. (2020). Research on insulator self-detonation defects based on multi-scale feature fusion Faster R-CNN[J]. *Hydropower Energy Sci.* 38 (11), 4.
- Zhu, M., Zhao, S., and Wang, J. (2021). Research on insulator detection method based on improved FasterR-CNN [J]. *Sci. Technol. Wind* 2213 (15), 012036. doi:10.1088/1742-6596/2213/1/012036
- Zou, D. (2022). *Deep learning-based insulator fault detection for transmission lines [D]*. Beijing: North China University of Electric Power.
- Zou, Z., Shi, Z., and Guo, Y. (2019). Object detection in 20 Years: A survey. arXiv.
- Zuo, G., Ma, L., and Xu, C. (2019). A cross-connected convolutional neural network-based insulator detection method[J]. *Power Syst. Autom.* 43 (4), 6. doi:10.7500/AEPS20180510002