Research Article

Insulators’ Identification and Missing Defect Detection in Aerial Images Based on Cascaded YOLO Models

Jingjing Liu,1 Chuanyang Liu,1,2 Yiquan Wu,2 Zuo Sun,1 and Huajie Xu1

1College of Mechanical and Electrical Engineering, Chizhou University, Chizhou 247000, China
2College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

Correspondence should be addressed to Chuanyang Liu; liuchuanyang608@nuaa.edu.cn

1.Introduction

With the continuous and rapid development of electric power in China, the total mileage of transmission lines also increases rapidly. The inspection of high-voltage transmission lines by an unmanned aerial vehicle (UAV) has attracted growing attentions in the power grid [1–3]. Conducting the regular inspection of key electrical equipment in high-voltage transmission lines is an effective way to ensure the safe and stable operation of the power grid [4], as shown in Figure 1(a). The aerial images captured by UAV consist of many different electrical equipment, e.g., insulator, antivibration hammers, and clamp are encircled by rectangular boxes with different colors. Specially, an insulator is an extremely important electrical equipment in high-voltage transmission lines, which plays the dual roles of mechanical support and electrical insulation [5, 6]. However, with the insulator’s perennial exposure to wind, sun, and rain, coupled with snow and ice coverage and other adverse weather effects, the missing defect of insulator in transmission lines is inevitable, as shown in Figure 1(b). The missing defect of the insulator is encircled by a blue ellipse. Statistically, most of the power grid accidents are caused by the missing defect of the insulator, and therefore, it is of great significance to automatically and timely perform insulators’ identification and their missing defect detection in high-voltage transmission lines [7, 8].

Some scholars have committed to research on insulators and their missing defect detection in transmission lines, and many valuable research results have been obtained though target detection algorithms [9–14]. Commonly, according to whether it is necessary to manually design features, target detection algorithms can be divided into the traditional image processing method and the method based on deep
The traditional method is composed of three parts, which are as follows: regions of interest (RoI) selection of different scales and widths, feature extraction, and target classification. Among them, the selection of RoI mostly uses the sliding window-based method. The manual designing method is the generally used manner for feature extraction (e.g., color feature, texture feature, scale invariant feature transform (SIFT), histogram of gradient (HOG), etc.). The support vector machine (SVM) and AdaBoost are the mainly used classifiers. Nevertheless, there are some disadvantages in the traditional method, such as the time-consuming nature and redundancy of the sliding window-based method, the manually designed feature is not robust to different environments, the computing ability is limited, and the processing ability of big data information is poor. Additionally, insulator aerial images are usually captured under complex backgrounds. Because of different shooting distances and angles, insulators and their missing defect are diverse, and they appear in different shapes. Consequently, in the practical applications, it is difficult for the traditional target detection method to recognize insulators and their missing defect in high-voltage transmission lines.

In the past decade, deep learning has made significant progress and become popular in the field of target detection, image classification, semantic segmentation, and so on [16–18]. Compared with the traditional target detection method, the method based on deep learning utilizes a nonlinear model to transform the low-level feature into high-level feature layer by layer and can automatically learn the useful features of the target, which have powerful learning and characterization capabilities, showing a strong advantage in feature extraction. Deep learning achieves the state-of-the-art performance in target detection by means of deep convolutional neural networks (DCNNs), which can represent the input image into high-dimensional feature space [19]. The development of deep learning has greatly promoted target detection, and target detection based on deep learning has become one of the research hotspots in the field of computer vision today. To improve the generalization performance of DCNNs, many excellent network structures of deep learning are proposed, particularly, LeNet, AlexNet, VGGNet, GoogleNet, ResNet, Densely Connected Convolutional Networks (DenseNet), and so on.

Currently, deep learning algorithms used for target detection are generally divided into two categories, one-stage algorithm based on regression, and two-stage algorithm based on region proposal. Specifically, the two-stage algorithm divides the target detection problem into two steps. First of all, the candidate regions are selected from the input image, and then the classification and regression are performed on the selected candidate regions. The representative algorithms of two-stage are faster regions with convolutional neural networks (faster R-CNN) [20–24], region-based fully convolutional networks (R-FCN) [25], mask R-CNN [26], and cascade R-CNN [27–29]. The application research of electrical equipment detection in transmission lines based on two-stage target detection algorithms is shown in Table 1. The two-stage algorithm performs well on large-scale datasets, however, such an algorithm needs a large number of calculations and runs slowly. Hence, it is difficult to achieve the effect of real-time detection.

To improve real-time detection performance while ensuring accuracy, one-stage algorithms represented by YOLO [30–32] and single shot multibox detector (SSD) [33] have emerged. One-stage algorithms are regression-based target detection networks, which use DCNNs to extract features and directly realize classification and regression to obtain the category probability and location information of the target. The two-stage algorithms are the representatives of accuracy, which can fully learn the features of the target because they obtain the candidate regions in advance, however, the structures of networks are complex, and the detection speed is slow. Hence, they are not suitable for real-time applications, while one-stage algorithms are the representatives of speed, which can meet the requirements of application scenarios for real-time detection. Consequently, one-stage

![Aerial images captured by UAV: (a) different electrical equipment of transmission lines in the image (e.g., insulator, antivibration hammers, clamp, etc.) and (b) missing defect of insulator. (a) Aerial image with different electrical equipment. (b) Aerial image with missing defect of insulator.](image-url)
algorithms have a high application research value for insulator identification, and many satisfactory detection results are obtained by the use of one-stage algorithms [34–42]. In the work of [38], a novel SSD algorithm is proposed to recognize and localize insulators in visible light images. To realize multisize insulator detection in the background interference with high precision and low time cost, auxiliary convolutional feature layers are employed to VGG-16 network to replace the fully connected layers. Experimental results show that the mean running time of per image is 0.03 s, which is much faster than that of faster R-CNN (0.11 s per image). In the work of Tian [39], a deep learning architecture is proposed to automatically detect insulator faults in aerial images. Firstly, the SSD algorithm is utilized to detect the position of insulators. Then, DenseNet is used to determine the category of the detected insulators. Finally, experiments are conducted on the testing dataset. The precision of insulator detection and classification have reached 0.95 and 0.98, respectively. Compared with the YOLO algorithms, the detection speed of the SSD algorithm needs to be further improved. To achieve multiple sizes of insulator detection in aerial images with complicated backgrounds, in the work of [40], three dense blocks are added to YOLOv3 to enhance feature propagation and reuse. The average precision of the modified YOLOv3 (YOLOv3-dense) model is 94.47%, and the average running time of per image can reach 8.5 ms, however, the modified model can only perform the localization detection of insulators, which is not used for the research of insulator defect detection. In [41], a deep neural network model is proposed to improve the accuracy and robustness of insulator detection in high-voltage transmission lines. The YOLOv3 model based on the backbone network of ResNet 50 is used to locate the position of insulator, and then the Grab-cut algorithm and the adaptive morphology method are used to perform insulator defect detection. Although the average running time of the YOLOv3 model can reach 0.02 s per image, it is difficult for the traditional image processing method to realize insulator defect detection in aerial images with diverse background interference. In [42], a cross stage partial dense YOLO (CSPD-YOLO) model is proposed for insulator defect detection, which achieves a good effect on the detection of the missing defect. Since the number of insulator defects is much lesser than that of the insulator, if the model is used for both insulators and their defect detection, it will lead to class imbalance [43].

In summary, as the two-stage algorithms are hard to train and difficult to deploy on the embedded devices, it can thus be concluded that one-stage algorithms have the potential to meet the requirements of real-time detection, and the application of YOLO models on high-voltage transmission lines automatic inspection have very considerable prospects. To perform both insulator identification and missing defect detection in aerial images from high-voltage transmission lines, based on the previous work [40], and inspired by the deep learning architectures [27–29], this paper proposed a cascaded YOLO models (improved YOLOv3-dense and YOLOv4-tiny) for insulators’ identification and their missing defect detection in aerial images. Specifically, improved YOLOv3-dense model is used to locate the insulators, and then the detected region of insulators is set as ROI area. Next, the improved YOLOv4-tiny model is employed to perform missing defect detection on ROI area.

The rest of this paper is organized as follows: Section 1 reports the related works of insulator identification and defect detection in aerial images. Section 2 describes the proposed method in details. Section 3 exhibits and discusses the experimental results and analysis of the proposed method. Finally, the conclusion and future work are shown in Section 4.

2. Materials and Methods

Since Hinton proposed the use of neural networks to automatically learn high-level feature in multimedia data, target detection based on deep learning has developed rapidly in the field of computer vision. Deep learning aims to locate the target of interest from the image, accurately determine the category of each target, and give the bounding box of each target. To obtain richer feature representation of target, on the one hand, large-scale image databases, such as ImageNet and COCO, are created. On the other hand, the commonly used backbone networks (e.g., VGGNet, GoogleNet, ResNet, etc.) are exploited to deepen DCNs, promoting the accuracy and execution efficiency of target

| Algorithm | Target detection | Detection effect |
|-----------|------------------|------------------|
| Faster R-CNN [20] | Insulators and their faults | Achieves a precision of 94% and a recall of 88% |
| Faster R-CNN [21] | Insulator detection | The average precision value reaches 0.818 using VGG-16 |
| Faster R-CNN [22] | 8 defects in transmission lines | The defects can be effectively and accurately identified |
| Faster R-CNN [23] | Insulator self-explosion defect | The identification rate reaches 89.0% |
| Faster R-CNN [24] | Insulator | The average precision at a level of 0.8 for 60 frames |
| R-FCN [25] | Cracked insulator detection | The average accuracy rate of 90.5% |
| Mask R-CNN [26] | Insulator identification and self-shattering detection | The mean average precision can be up to 1 for single-target images and 0.948 for multitarget images |
| LIN + DDN [27] | Insulator defect detection | Defect detection precision and recall are 0.91 and 0.96 |
| Faster R-CNN + FCN [28] | Identification of insulator explosion | The faulted insulators can be effectively detected in highly cluttered images |
| Faster R-CNN + Res-UNet [29] | Insulator defect identification | The precision and recall are 91.9% and 95.7% |
detection greatly. Recently, methods based on deep learning are widely used in the fields of video surveillance, intelligent transportation, the automatic driving of vehicles, pedestrian detection, face identification, robot navigation, and so on. Inspired by these state-of-the-art works, it is worth studying how to use deep learning algorithms for insulators’ identification and their missing defect detection in UAV aerial images.

Cascaded deep learning models are designed to automatically detect insulators and their missing defects in complex aerial images. The improved YOLOv3-dense model and YOLOv4-tiny model are cascaded, which transforms insulator identification and missing defect detection to a two-level detection problem, directly uses improved YOLOv3-dense model to locate the position of insulator, and then improved YOLOv4-tiny model is used to perform the defect detection on the insulator. The entire architecture of the proposed method is shown in Figure 2. Firstly, insulator aerial images captured by UAV are used to construct the “CCIN_detection” (Chinese Composite Insulator) dataset [40], which is divided into the training dataset, testing dataset, and “InSF-detection” (Insulator Fault, InsSF) dataset [42]. Secondly, the training dataset is utilized to train the improved YOLOv3-dense model for locating the insulator position, and the “InSF-detection” dataset is used to train the improved YOLOv4-tiny model for missing defect detection. The final models for insulator recognition and missing defect detecting are obtained after training. Finally, to verify the effectiveness of the proposed method, the final detection models are cascaded to perform insulator identification and missing defect detection on the testing dataset.

2.1. Insulators’ Identification. Insulator aerial images captured by UAV usually include complex backgrounds, such as river, farmland, power tower, building, and so on. In addition, since the characteristics are of multitype and multiscale, the sizes of insulator are extremely diverse, resulting in difficult-to-recognize insulators in aerial images. Therefore, how to effectively extract and locate insulators in aerial images with complex background is very important. In our previous work [40], the YOLOv3-dense model (three dense blocks are adopted to the feature extraction network of the YOLOv3 model) is proposed to recognize the multisize of insulators amid diverse background interference. Experimental results conducted on the testing set of CCIN_detection show that the average precision (AP) value of the YOLOv3-dense model can reach 94.47%, which is higher than that of the YOLOv2 model (83.43%) and YOLOv3 model (90.31%). Although the experimental results verify that our proposed YOLOv3-dense model can effectively recognize insulators from the complex backgrounds, we decided to explore a new DCNN model, which is more accurate than the previous YOLOv3-dense model. To improve the accuracy of insulator identification, this section mainly focuses on the improvement of the feature extraction network and feature fusion based on YOLOv3-dense model. The entire structure of the proposed model is shown in Figure 3, which is made of feature extraction network, multiscale prediction network, and detection network.

2.1.1. Feature Extraction Network of the Proposed Model. The depth of DCNNs is very important to the detection performance of the deep learning model. The deeper network can extract high-efficiency features for targets identification, however, the detection accuracy of the model will quickly reach saturation as the network depth increases, and it even begins to decline rapidly. To solve the problem of network degradation, residual network (ResNet) is proposed by He et al. [44], so that the network with hundreds of layers can be trained. The main expression of ResNet is defined in formula (1).

\[
x_l = H_l(x_{l-1}) + x_{l-1},
\]

where \(l\) denotes the layer of DCNNs, \(x_l\) indicates the output of layer \(l\), and \(H_l\) represents the nonlinear transformation of \(x_{l-1}\). In the ResNet module, the Conv \((1 \times 1)\) and Conv \((3 \times 3)\) are mainly employed. Conv \((1 \times 1)\) is used for compression feature representation, and Conv \((3 \times 3)\) is utilized for feature extraction.

To effectively reduce the parameters of the network model while keeping the low-level feature in the high-level layers as much as possible, the architecture of DenseNet is proposed by Huang et al. [45], which uses the dense connection between the layers to achieve information integration, avoiding the loss of feature information and the disappearance of gradients, so that the network can learn more features between different levels and improve the detection accuracy to a certain extent. The feed-forward manner is adopted to DenseNet to connect each layer to every other layer. In other words, the layer \(l\) receives all the features of the previous \(l - 1\) layers as an input. The formula is defined as follows:

\[
x_l = H_l(\{x_0, x_1, \ldots, x_{l-1}\}),
\]

where \([x_0, x_1, \ldots, x_{l-1}]\) denotes the output feature maps of layers \([x_0, x_1, \ldots, x_{l-1}]\) spliced in series. Moreover, the transform function \(H_l\) is a nonlinear function BN-ReLU-Conv \((1 \times 1)\)-BN-ReLU-Conv \((3 \times 3)\), which is a combination of batch normalisation (BN), rectified linear unit (ReLU), and convolutional (Conv).

To further improve the feature extraction network of YOLOv3-dense, CSPNet is adopted to the YOLOv3-dense model in this paper. CSPNet is a new type of backbone network proposed by Wang et al. [46] in 2020, which can be used to enhance the learning ability of DCNNs, maintain accuracy while being lightweight, reduce the computational bottlenecks, and reduce the cost of memory. In the structure of CSPNet, the feature maps of input layer are split into two parts: the first part is actually to yield a large residual edge, which is directly connected to the end after one convolutional operation. The latter is the main feature extraction part, which goes through a residual block, i.e., the number of channels is adjusted by Conv \((1 \times 1)\). Then, feature extraction is enhanced by Conv \((3 \times 3)\). The output is concatenated with
the small residual edge, and finally, the number of channels is adjusted to be the same as the first part after the Conv $(1 \times 1)$ operation. The feature extraction network of the proposed model is shown in Table 2.

As shown in Table 2, the feature extraction network of the proposed model can be divided into six parts, as follows: a Conv $(3 \times 3)$ is used for feature $(416 \times 416)$ extraction in the first part. A cross-stage partial convolution and one residual module are used for feature $(208 \times 208)$ extraction in the second part. A cross-stage partial convolution and two residual modules are utilized for feature $(104 \times 104)$ extraction in the third part. From the forth part to the sixth part, a cross-stage partial convolution, four dense modules, and four residual modules are employed for effective feature $(52 \times 52, 26 \times 26, 13 \times 13)$ extraction, as shown in Figure 4. Specifically, the input feature maps are $(104 \times 104 \times 128)$ in the fourth part—Conv $(256 \times 1 \times 1)$, Conv $(256 \times 3 \times 3/2)$, four dense modules, four residual modules, Conv $(128 \times 1 \times 1)$, and Conv $(256 \times 1 \times 1)$ are used to extract feature $(52 \times 52)$. Firstly, Conv $(256 \times 3 \times 3/2)$ is used to compress image dimensions, and the feature maps $(52 \times 52 \times 256)$ are obtained. Secondly, one part of feature maps $(52 \times 52 \times 256)$ through Conv $(256 \times 1 \times 1)$ to yield a large residual edge, and the other part of feature maps $(52 \times 52 \times 256)$ is the main feature extraction part, which goes through four dense modules and four residual
modules for feature extraction. During the process of feature extraction, the number of channels is adjusted to 128 through Conv \((128 \times 1 \times 1)\), and BN-ReLU-Conv \((32 \times 1 \times 1)\)-BN-ReLU-Conv \((32 \times 3 \times 3)\) are applied to each dense module for feature extraction. The four residual modules are used to continue extracting feature. Finally, the feature maps \((52 \times 52)\) generated from the two parts are concatenated, and then the number of channels is adjusted to 256 through Conv \((256 \times 1 \times 1)\) to obtain the output feature maps \((26 \times 26 \times 512)\). In the sixth part—Conv \((1024 \times 1 \times 1)\), Conv \((1024 \times 3 \times 3/2)\), four dense modules, four residual modules, Conv \((512 \times 1 \times 1)\), and Conv \((1024 \times 1 \times 1)\) are used to extract feature \((13 \times 13)\). BN-ReLU-Conv \((128 \times 1 \times 1)\)-BN-ReLU-Conv \((128 \times 3 \times 3)\) are applied to each dense module for feature extraction. The feature maps \((13 \times 13)\) generated from the two parts are concatenated, and then the number of channels is adjusted to 1024 through Conv \((1024 \times 1 \times 1)\) to yield the output feature maps \((13 \times 13 \times 1024)\).

### 2.1.2. Multiscale Pooling Module

The pooling operation is indispensable for DCNNs, which can reduce the size of images, while preserving the main information of the image as much as possible, so that the network model can focus more on the main feature. Spatial pyramid pooling (SPP) is a method proposed by He et al. [47] to solve the problem of images with different sizes input to the DCNNs, generating a

| Cross-stage partial | Type          | Filters | Size          | Output             |
|---------------------|---------------|---------|---------------|--------------------|
|                     | Convolutional | 32      | \(3 \times 3/1\) | \(416 \times 416 \times 32\) |
|                     | Convolutional | 64      | \(3 \times 3/2\) |                     |
| Conv \((64 \times 1 \times 1)\) | 1× Convolutional | 32      | \(1 \times 1/1\) | \(208 \times 208 \times 64\) |
|                     | Convolutional | 64      | \(3 \times 3/1\) |                     |
|                     | Residual      | 64      | \(1 \times 1/1\) |                     |
|                     | Convolutional | 128     | \(3 \times 3/2\) |                     |
|                     | Convolutional | 128     | \(1 \times 1/1\) |                     |
| Conv \((128 \times 1 \times 1)\) | 2× Convolutional | 64      | \(1 \times 1/1\) | \(104 \times 104 \times 128\) |
|                     | Convolutional | 128     | \(3 \times 3/1\) |                     |
|                     | Residual      | 128     | \(1 \times 1/1\) |                     |
|                     | Convolutional | 256     | \(3 \times 3/2\) |                     |
|                     | Convolutional | 128     | \(1 \times 1/1\) |                     |
| Conv \((256 \times 1 \times 1)\) | 4× Convolutional | 32      | \(1 \times 1/1\) | \(52 \times 52 \times 256\) |
|                     | Convolutional | 32      | \(3 \times 3/1\) |                     |
|                     | DenseNet      | 128     | \(1 \times 1/1\) |                     |
|                     | Convolutional | 256     | \(3 \times 3/1\) |                     |
|                     | Residual      | 256     | \(1 \times 1/1\) |                     |
|                     | Convolutional | 512     | \(3 \times 3/2\) |                     |
| Conv \((512 \times 1 \times 1)\) | 4× Convolutional | 256     | \(1 \times 1/1\) | \(26 \times 26 \times 512\) |
|                     | Convolutional | 64      | \(1 \times 1/1\) |                     |
|                     | Convolutional | 64      | \(3 \times 3/1\) |                     |
|                     | DenseNet      | 128     | \(1 \times 1/1\) |                     |
|                     | Convolutional | 256     | \(3 \times 3/1\) |                     |
|                     | Residual      | 512     | \(1 \times 1/1\) |                     |
|                     | Convolutional | 1024    | \(3 \times 3/2\) |                     |
| Conv \((1024 \times 1 \times 1)\) | 4× Convolutional | 256     | \(1 \times 1/1\) | \(13 \times 13 \times 1024\) |
|                     | Convolutional | 128     | \(1 \times 1/1\) |                     |
|                     | Convolutional | 128     | \(3 \times 3/1\) |                     |
|                     | DenseNet      | 512     | \(1 \times 1/1\) |                     |
|                     | Convolutional | 1024    | \(3 \times 3/3\) |                     |
|                     | Residual      | 1024    | \(1 \times 1/1\) |                     |
fixed-size feature representation without considering the size or proportion of the input image. Different from the purpose of work [47], a structure of SPPNet is employed in this paper, aiming to obtain multiscale local features information and fuse them with global feature information to enrich the expressive ability of the feature maps, thereby improving the accuracy of target prediction.

To further obtain more local features, a multiscale pooling module is used in the last feature layer of the feature extraction network, as shown in Figure 5. Three max-pooling layers of different kernel sizes (5 × 5, 9 × 9, 13 × 13) are used to obtain multiscale local feature maps, and then the local feature maps are concatenated with the input global feature maps. By the use of the multiscale pooling module, the receptive field of the last feature layer can be expanded. In addition, the richer local features information can be obtained, and the sensitivity of the target with different scales can be improved.

2.1.3. Multiscale Prediction Network. In the previous YOLOv3-dense model, the effective feature maps (52 × 52, 26 × 26, 13 × 13) that corresponded to different levels of feature information are selected for feature fusion. As the span between these feature maps is very large, after multiple convolution operations, the feature at each level retains less feature information from other levels. As the number of network layers continues to deepen, the feature extraction of semantic information becomes better, however, with the enhancement of semantic information, the loss of detailed information (e.g., shape feature, color feature, texture feature, etc.) brings many negative effects on insulator localization, resulting in it being impossible to accurately locate the position of insulators. To enhance the feature representation and further realize feature reuse, the top-down and bottom-up fusion strategy [48] is employed to the multiscale prediction network. The structure of multiscale prediction network is shown in Figure 6.

As shown in Figure 6, the process of multiscale prediction is as follows: firstly, the original image is subjected to feature extraction through the forward propagation of DCNNs. The size of the feature layers is gradually reduced, and the effective feature maps (C3, C4, and C5) are extracted. Secondly, convolutional operation and SPP are performed on feature maps C5 to generate feature maps P3, and then the feature maps P3 are fused with feature maps C4 via convolutional operation and up-sampling to obtain feature maps P2, while the feature maps P2 are fused with feature maps C3 via convolutional operation and up-sampling to yield feature maps P1. Finally, the convolutional operation is performed on feature maps P1 to generate feature maps N1, and then the feature maps N1 are fused with the feature maps P2 via convolutional operation and down-sampling to obtain feature maps N2, while the feature maps N2 are fused with the feature maps P3 via convolutional operation and down-sampling to generate feature maps N3. The feature maps N1, N2, and N3 are connected to the residual modules for feature (52 × 52, 26 × 26, 13 × 13) prediction, respectively. The feature maps from P3 to P1 is a top-down process. The high-level feature that contains more semantic information is fused with low-level feature through up-sampling, and the feature maps from N1 to N3 is a bottom-up process. The detailed information of low-level feature and the semantic information of high-level feature are retained for more accurate locating prediction.
2.2. **Missing Defect Detection.** In the process of transmission lines inspection based on UAV, it is commonly faced with complex scenes and changeable climate, which makes it difficult to inspect high-voltage transmission lines. Furthermore, there are still safety distance restrictions for UAV to fly along high-voltage transmission lines, resulting in the electrical equipment being captured by UAV accounts for a very small proportion to the whole image. Since most of insulator missing defect area is relatively small, the information of the defect area will be lost after multiple convolutional operation. As a result, the one-stage algorithm will not be able to effectively detect the position of insulator missing defect. In this work, cascaded deep learning models are employed for insulator identification and missing defect detection. Firstly, the first detection model (improved YOLOv3-dense model) detailed in section 2.1 is adopted to locate the position of insulators in aerial images. As a result, the one-stage algorithm will not be able to effectively detect the position of insulator missing defect. In this work, cascaded deep learning models are employed for insulator identification and missing defect detection. Firstly, the first detection model (improved YOLOv3-dense model) detailed in section 2.1 is adopted to locate the position of insulators in aerial images. Then, the localization of insulator detected by the proposed model is set as the RoI area. Finally, the second detection model (improved YOLOv4-tiny) is used to perform missing defect detection on the RoI area.

YOLOv4-tiny network is an abbreviated version of YOLOv4 model, which uses a more lightweight backbone network to extract image features, simplifies the feature fusion part of the neck, reduces the output branch of the head, and improves the detection speed of the network. More importantly, the memory usage of the final weights for the YOLOv4-tiny model is less than 30 MB, so that it is suitable for running on mobile or embedded devices for real-time detection. To improve the detection accuracy of insulator missing defect, we improve YOLOv4-tiny on the multiscale prediction network. The feature (52 × 52) detection head (in the red rectangular box) is added to the prediction network, and the entire structure of improved YOLOv4-tiny model is shown in Figure 7.

### 3. Experiments’ Results and Discussion

3.1. **Experimental Environment.** In this study, we conduct training and testing by the use of the deep learning framework Dark-net [49]. A server equipped with hardware and software is constructed to evaluate the proposed method. The basic configuration of the hardware platform is as follows: an Intel (R) Core (TM) i9-9900K CPU and an NVIDIA GeForce GTX 3080 GPU. The basic configuration of the software platform is as follows: Windows 10 as the
basic operating system, Opencv3.4.0 and Visual Studio 2017 as the graphics processing tools, CUDA 11.1 as parallel computing architecture, and cuDNN 8.0.5 as DCNNs acceleration library. The detail information of experimental environment is shown in Table 3.

### 3.2. Dataset Preparation

There is no known existing image dataset toward the missing defect estimation of the insulator. Hence, data collection is a significant part of the effort for this paper. For a fair comparison with YOLOv3-dense model, the proposed model described in section 2.1 is trained on the CCIN_detection dataset. First of all, we collect 5000 aerial images in total, which are more diverse and contain more common aerial scenes compared with the images in CPLID (Chinese Power Line Insulator Dataset) [27]. Next, all the images are resized to the same size of 416×416 pixels, and the LabelImg tool is utilized to label the ground-truth for all the insulator images. Finally, we randomly select 3000 images as a training set, and the other 2000 images are assigned to be a testing set.

Since the number of insulator missing defect is usually less than that of insulators in aerial images and the position of the missing defect is uncertain, it is impossible to get enough missing defect samples of real scenes for conducting insulator missing defect detection. To address the above issues, the image processing tool of Photoshop is used to obtain simulated missing defect samples. Ultimately, the InSF-detection dataset with 1331 aerial images is created, in which 809 images are set to be a training set, while the other 522 images are set to be a testing set. The relevant information of two datasets is shown in Table 4.

### 3.3. Quantitative and Qualitative Analysis

To evaluate the effectiveness of the proposed model described in Section 2.1, extensive experiments are conducted on the proposed model and the compared models (YOLOv3 and YOLOv3-dense).

For a fair comparison, the proposed model, YOLOv3, and YOLOv3-dense are trained and tested on the same dataset CCIN_detection. During the process of training, the number of iterations is set to be 38,000, the initial value of the learning rate is set as 0.001, decay = 0.005, momentum = 0.9, saturation = 1.5, hue = 0.1, exposure = 1.5, batch size = 64, and subdivision = 16, respectively.

In addition, average precision (AP), precision (P), and recall (R) are used to verify the effects of the proposed model quantitatively. The true positive (TP), false positive (FP), true negative (TN), and false negative (FN) are four common types used to evaluate the binary classification model. The definitions of TP, FP, TN, and FN are shown in Table 5. Among them, TP represents the number of samples where the detected category is consistent with the real target category, FP indicates the number of samples in which the detected target category is inconsistent with the real target category, TN represents the number of negative samples that is correctly classified, and FN indicates the number of samples in which the real target exists but has not been detected.

**Precision (P)** can be defined as the proportion of all correctly predicted targets to all the predicted targets, as shown in formula (3).

\[
\text{Precision}(P) = \frac{TP}{TP + FP}
\]
Recall (R) can be defined as the proportion of all correctly predicted targets to all the targets that should be predicted, as shown in formula (4).

\[
Recall (R) = \frac{TP}{TP + FN}
\]  

(4)

The precision-recall (P-R) curve is usually used to evaluate the performance of deep learning models, which can be obtained by taking the recall value as the horizontal axis and the precision value as the vertical axis. The AP value can be calculated by the area under the P-R curve, as shown in formula (5).

\[
AP = \int_{0}^{1} P(R) \, dR.
\]  

(5)

3.3.1. Insulators’ Identification Based on the Proposed YOLO Model. To evaluate the effect of insulators’ identification, the proposed model and the compared models (YOLOv3 and YOLOv3-dense) are trained and then tested. The experimental effects (AP, precision, and recall) of three models for insulators’ identification are listed in Table 6. Specifically, the AP values of three insulator identification models are as follows: YOLOv3 (90.3%), YOLOv3-dense (94.4%), and the proposed model (96.3%). It is found that the AP value of the proposed model is 6% higher than that of YOLOv3 and 1.9% higher than that of YOLOv3-dense, which demonstrates that the proposed detection model is more accurate than YOLOv3 and YOLOv3-dense. The precision values of the three insulator detection models are as follows: YOLOv3 (90%), YOLOv3-dense (94%), and the proposed model (96%). The precision value of the proposed detection model is 6% higher than that of YOLOv3 and 2% higher than that of YOLOv3-dense. The recall values of the three insulator detection models are as follows: YOLOv3 (91%), YOLOv3-dense (96%), and the proposed model (97%). The recall value of the proposed detection model is 6% higher than that of YOLOv3 and 1% higher than that of YOLOv3-dense. It can be concluded that the proposed model is more advantageous than YOLOv3 and YOLOv3-dense for insulators’ identification.

To further analyze the impact of the improved method proposed in this paper on the YOLOv3 model, ablation experiments are conducted. The algorithm in this paper is cut into five groups for training, respectively. The first group is the original YOLOv3 model, the second group is the feature extraction network structure of YOLOv3 changed to CSPNet, the third group is the multiscale pooling module (SPPNet) added to the previous group, the fourth group is the addition of multiscale prediction network (PANet) on the basis of the previous group, and the fifth group is the residual module used on the basis of the previous group, i.e., the fifth group is the algorithm proposed in this paper. Experimental results of the five groups are shown in Table 7.

It can be obtained from Table 7 for the first group of the experiment. The AP value of the original YOLOv3 model on insulators’ identification is 90.3%, and the FPS (frames per second) is 125. For the second group of the experiment, because of the introduction of CSPNet, the AP value of the insulators’ identification increases by 3.5%, and at the same time, the increase of FPS is 5. The cross-stage partial network enhances the learning ability of the convolutional network, eliminates most of the computational bottleneck structure, and reduces the memory consumption, thus improving the accuracy of detection and inference speed. For the third group of experiment, since the improved spatial pyramid pooling structure is added on the basis of the second group, the AP value increases by 1%, which shows that the improved spatial pyramid pooling structure is meaningful for improving the accuracy of model detection. For the fourth group of the experiment, since the path aggregation network is added on the basis of the third group, the AP on insulators’ identification can improve to a certain extent. It is because a bottom-up fusion path is added to the three effective feature layers, which further improves the detection effect of the prediction network. For the fifth group of experiment, the residual module is introduced on the basis of the fourth group. Compared with the fourth group, the AP value increases by 0.8%, especially compared with the original YOLOv3 model, which has a significant improvement as a whole. Meanwhile, the proposed model also achieves better real-time effect, and the FPS increases by 10 compared with YOLOv3. In summary, the improvement strategies proposed in this paper for YOLOv3 model are meaningful for improving the effects of insulators’ identification in complex scenes.

Insulator images captured by UAV usually contain complex backgrounds, e.g., river, farmland, power tower, building, sky, and so on. Moreover, the sizes of insulators in

---

**Table 4: The relevant information of two datasets.**

| Dataset         | Number of images | Training set | Testing set | Image size |
|-----------------|------------------|--------------|-------------|------------|
| CCIN_detection  | 5000             | 3000         | 2000        | 416 × 416  |
| InSF-detection  | 1331             | 809          | 522         | 416 × 416  |

**Table 5: The definition of TP, FP, TN, and FN.**

| Ground truth | Prediction result | Definition |
|--------------|-------------------|------------|
| Positive (1) | Positive (1)      | TP         |
| Positive (1) | Negative (0)      | FN         |
| Negative (0) | Positive (1)      | FP         |
| Negative (0) | Negative (0)      | TN         |

**Table 6: The experimental effects of three insulator identification models.**

| Networks     | AP (%) | Precision (%) | Recall (%) |
|--------------|--------|---------------|------------|
| YOLOv3       | 90.3   | 90            | 91         |
| YOLOv3-dense | 94.4   | 94            | 96         |
| Proposed model | 96.3  | 96            | 97         |
insulator missing defect prediction are listed in Table 8. Concretely, the AP values of the three network models are as follows: YOLOv3-tiny (89.6%), YOLOv3 (93.3%), and the improved YOLOv4-tiny (89%). It is found that the AP value of the improved YOLOv4-tiny is just a little lower than that of YOLOv3. The memory usages of the three network models are as follows: YOLOv3-tiny (33 MB), YOLOv3 (240 MB), and the improved YOLOv4-tiny (24 MB), and the running times of the three network models are as follows: YOLOv3-tiny (4.5 ms), YOLOv3 (10 ms), and the improved YOLOv4-tiny (4 ms). It can be observed that improved YOLOv4-tiny not only takes up less memory usage but also has faster detection speed. Consequently, considering the AP values, memory usages, and running times, it can be concluded that the improved YOLOv4-tiny model is suitable to deploy on embedded devices for insulator defect detection.

Experimental results with different scenes conducted by improved YOLOv4-tiny model are shown in Figure 9. Figures 9(a)–9(d) demonstrate the experimental scenes with the background of river, farmland, power tower, and sky, respectively. From the figures 9(a)–9(d), it is found that all the insulators with a single-defect or multiple-defect are accurately detected in aerial images with different background interference by the improved YOLOv4-tiny model.

### Table 7: Comparison of ablation experimental results.

| Methods          | CSPNet | SPPNet | PANet | Residual module | AP (%) | FPS |
|------------------|--------|--------|-------|-----------------|--------|-----|
| YOLOv3 (1)       |        |        |       |                 | 90.3   | 125 |
| YOLOv3 (2)       | √      |        |       |                 | 93.8   | 130 |
| YOLOv3 (3)       | √      | √      |       |                 | 94.8   | 128 |
| YOLOv3 (4)       | √      | √      | √     |                 | 95.5   | 132 |
| YOLOv3 (5)       | √      | √      | √     |                 | 96.3   | 135 |

#### 3.3.2. Insulator Missing Defect Detection Based on Improved YOLOv4-Tiny

To evaluate the effect of insulator missing defect detection, YOLOv3-tiny, YOLOv3, and improved YOLOv4-tiny are trained and tested on the dataset InSF-detection. The experimental effects (AP, precision, recall, memory usage, and running time) of the three models for insulator missing defect prediction are listed in Table 8. Specifically, the AP values of the three network models are as follows: YOLOv3-tiny (89.6%), YOLOv3 (93.3%), and the improved YOLOv4-tiny (92.9%). The Recall values of the three network models are as follows: YOLOv3-tiny (94%), YOLOv3 (94%), and the improved YOLOv4-tiny (94%). The Recall values of the three network models are as follows: YOLOv3-tiny (89%), YOLOv3 (94%), and the improved YOLOv4-tiny (89%). It is found that the AP value of the improved YOLOv4-tiny is just a little lower than that of YOLOv3. The memory usages of the three network models are as follows: YOLOv3-tiny (33 MB), YOLOv3 (240 MB), and the improved YOLOv4-tiny (24 MB), and the running times of the three network models are as follows: YOLOv3-tiny (4.5 ms), YOLOv3 (10 ms), and the improved YOLOv4-tiny (4 ms). It can be observed that improved YOLOv4-tiny not only takes up less memory usage but also has faster detection speed. Consequently, considering the AP values, memory usages, and running times, it can be concluded that the improved YOLOv4-tiny model is suitable to deploy on embedded devices for insulator defect detection.

Experimental results with different scenes conducted by improved YOLOv4-tiny model are shown in Figure 9. Figures 9(a)–9(d) demonstrate the experimental scenes with the background of river, farmland, power tower, and sky, respectively. From the figures 9(a)–9(d), it is found that all the insulators with a single-defect or multiple-defect are accurately detected in aerial images with different background interference by the improved YOLOv4-tiny model.

#### 3.3.3. The Cascaded YOLO Models for Insulators Detection

To verify the accuracy and robustness of the proposed method, improved YOLOv3-dense model and YOLOv4-tiny model are cascaded to perform insulators’ identification and their missing defect detection. The cascaded YOLO models, faster RCNN, and SSD are tested on the testing dataset of “CCIN_detection,” and YOLOv4, models in Literature [42, 50], faster RCNN, SSD, our proposed models are used for insulator missing defect detection on the testing dataset of “InSF-detection.” The AP, precision, recall, and FPS values of the six models for insulator missing defect prediction are listed in Table 9. Concretely, the AP values of the six models are as follows: YOLOv4 (96.38%), model in [50] (96.5%), model in [42] (98.18%), Faster RCNN (93.2%), SSD (88.1%), and the cascaded YOLO models (98.4%). The precision values of the six models are as follows: YOLOv4 (98%), model in [50] (98%), model in [42] (99%), Faster RCNN (94%), SSD (85%), and the cascaded YOLO models (99%). The recall values of the six models are as follows: YOLOv4 (95%), model in Literature [50] (95%), model in [42] (98%), Faster RCNN (92%), SSD (89%), and the cascaded YOLO models (98%). The FPS values of the six models are as follows: YOLOv4 (100), model in [50] (100), model in [42] (91), Faster RCNN (35), SSD (78), and the cascaded YOLO models (106), respectively. Since insulator defect detection can be regarded as small target detection and the AP value of our proposed model is 98.4%, which is 5.2% higher than that of faster RCNN and 10.2% higher than that of SSD, our proposed model has good ability for small object detection. Simultaneously, the FPS value of our proposed model can reach 106, which is the biggest in the six models. Accordingly, it can be found that the cascaded YOLO models can be more advantageous than the faster RCNN,
SSD, YOLOv4, and models in Literature [42, 50] in insulator missing defect prediction.

Some typical images with different scenes are selected to exhibit the visualization performance of the cascaded YOLO models, as shown in Figure 10. The experimental scenes with backgrounds of river, farmland, and power tower were exhibited in figures 10(a)–10(c), the experimental scene with the insulator’s color being similar to the background is shown in Figure 10(d), and the experimental scene with the occlusion of insulator strings is demonstrated in

### Table 8: The experimental effects of three insulator missing defect detection models.

| Networks           | AP (%) | Precision (%) | Recall (%) | Memory usage (MB) | Running time (ms) |
|--------------------|--------|---------------|------------|-------------------|------------------|
| YOLOv3-tiny        | 89.6   | 92            | 89         | 33                | 4.5              |
| YOLOv3             | 93.3   | 94            | 94         | 240               | 10               |
| Improved YOLOv4-tiny | 92.9   | 94            | 89         | 24                | 4                |

**Figure 8:** Experimental results with different scenes conducted by the proposed model (from the first row to the fifth row show experimental scenes with the background of river, farmland, power tower, building, and sky, respectively). (a) Experimental scene with the background of a river, (b) experimental scene with the background of farmland, (c) experimental scene with the background of a power tower, (d) experimental scene with the background of a building, and (e) experimental scene with the background of the sky.

**Figure 9:** Experimental results with different scenes conducted by improved YOLOv4-tiny. (a) River background, (b) farmland background, (c) power tower background, and (d) sky background.
Figure 10(e). Although the backgrounds are complex, the insulator’s color is similar to the background, and part of the insulator strings are occluded. All the insulators and their defects are accurately detected by the cascaded YOLO models. Consequently, it may be concluded that the proposed method (improved YOLOv3-dense and YOLOv4-tiny are cascaded) can obtain good performance in insulators’ identification and their missing defect detection in aerial images with complex backgrounds.

4. Conclusions

This study presented cascaded YOLO models for insulators’ identification and their missing defect detection in aerial images from high-voltage transmission lines. First of all, “CCIN_detection” dataset, including 5000 insulator images and “InSF-detection” dataset consisting of 1331 insulator defect images were created. After that, to improve the accuracy of insulator detection from complex background interference, the feature extraction network and multiscale prediction network were improved in our previous YOLOv3-dense model. Moreover, the detected area of insulator was set as RoI area, and improved YOLOv4-tiny model with three detection heads was used to perform missing defect detection on the RoI area. Finally, the improved YOLOv3-dense and YOLOv4-tiny models were carefully trained, and the final models were cascaded to conduct insulator identification and defect detection on the testing set. Extensive experiments were performed on the cascaded YOLO models and the compared models. The experimental results showed that the AP values of YOLOv3-dense, YOLOv3, and the improved YOLOv3-dense were

| Networks        | AP (%) | Precision (%) | Recall (%) | FPS |
|-----------------|--------|---------------|------------|-----|
| YOLOv4          | 96.38  | 98            | 95         | 100 |
| Literature [50] | 96.5   | 98            | 95         | 100 |
| Literature [42] | 98.18  | 99            | 98         | 91  |
| Faster RCNN     | 93.2   | 94            | 92         | 35  |
| SSD             | 88.1   | 85            | 89         | 78  |
| Our proposed models | 98.40  | 99            | 98         | 106 |

Figure 10: Experimental results with different scenes conducted by the cascaded models. (a) Experimental scene with a river in the background. (b) Experimental scene with a farmland in the background. (c) Experimental scene with a power tower in the background. (d) Experimental scene with the insulator’s color being similar to the background. (e) Experimental scene with the occlusion of insulator strings.
94.4%, 90.3%, and 96.3%, respectively, proving the proposed model was superior to YOLOv3 and YOLOv3-tiny models. The AP value of the improved YOLOv4-tiny (92.6%) model was slightly lower than that of YOLOv3 (93.3%) model. More importantly, the memory usage of the improved YOLOv4-tiny (24 MB) model was much less than that of YOLOv3 (240 MB) and YOLOv3-tiny (33 MB) models. The AP and FPS values of cascaded YOLO models for insulator missing defect detection could reach 98.40% and 106, which are higher than those of the compared models. Therefore, it can be concluded that the cascaded models can obtain good performance in insulators’ identification and their missing defect detection in aerial images and have the potential of real-time inspection for high-voltage transmission lines.

For a future study, the samples of insulators and their missing defect should be collected and added to datasets “CCIN_detection” and “InSF-detection.” In addition, the proposed models should be adjusted to adapt to more types of insulator defect detection.

Data Availability
No data were used to support this study.

Conflicts of Interest
The authors declare that they have no conflicts of interest.

Authors’ Contributions
J.L. took part in writing and finishing the experiments in this paper. Y. W. contributed for methodology. C. L. was responsible for experiments. H. X. took part in data collection and analysis. Z. S. was responsible for labeling the ground truth for each image in our dataset. All authors have read and agreed to the published version of the manuscript.

Acknowledgments
The authors thank Yiguan Wu for his guidance. This research was funded by the National Nature Science Founding of China, under grant 61573183, and Major Natural Research Project of Anhui Provincial, under grant KJ2019ZD63.

References
[1] V. N. Nguyen, R. Jenssen, and D. Roverso, “Automatic autonomous vision-based power line inspection: a review of current status and the potential role of deep learning,” International Journal of Electrical Power & Energy Systems, vol. 99, pp. 107–120, 2018.
[2] J. Han, Z. Yang, H. Xu et al., “Search like an eagle: a cascaded model for insulator missing faults detection in aerial images,” Energies, vol. 13, no. 3, p. 713, 2020.
[3] Y. Ma, Q. Li, L. Chu, Y. Zhou, and C. Xu, “Real-time detection and spatial localization of insulators for UAV inspection based on binocular stereo vision,” Remote Sensing, vol. 13, no. 2, p. 230, 2021.
[4] L. Yang, J. Fan, Y. Liu, E. Li, J. Peng, and Z. Liang, “A review on state-of-the-art power line inspection techniques,” IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 12, pp. 9350–9365, 2020.
[5] Y. Liu, S. Pei, W. Fu, K. Zhang, X. Ji, and Z. Yin, “The discrimination method as applied to a deteriorated porcelain insulator used in transmission lines on the basis of a convolution neural network,” IEEE Transactions on Dielectrics and Electrical Insulation, vol. 24, no. 6, Article ID 060840, pp. 3559–3566, 2017.
[6] D. Zuo, H. Hu, R. Qian, and Z. Liu, “An insulator defect detection algorithm based on computer vision,” in Proceedings of the 2017 IEEE International Conference on Information and Automation (ICIA), pp. 361–365, Macau, China, July 2017.
[7] Y. Zhai, D. Wang, M. Zhang, J. Wang, and F. Guo, “Fault detection of insulator based on saliency and adaptive morphology,” Multimedia Tools and Applications, vol. 76, no. 9, pp. 12051–12064, 2017.
[8] Y. Zhai, R. Chen, Q. Yang, X. Li, and Z. Zhao, “Insulator fault detection based on spatial morphological features of aerial images,” IEEE Access, vol. 6 Article ID 2846293, pp. 35316–35326, 2018.
[9] M. Wang, Y. Du, and Z. Zhang, “Study on power transmission lines inspection using unmanned aerial vehicle and image identification of insulator defect,” Journal of Electronic Measurement and Instrument, vol. 29, pp. 1862–1869, 2015.
[10] S. Liao and J. An, “A robust insulator detection algorithm based on local features and spatial orders for aerial images,” IEEE Geoscience and Remote Sensing Letters, vol. 12, pp. 963–967, 2017.
[11] Y. Jiang, J. Han, J. Ding, H. Fu, Y. Wang, and W. Cao, “The identification and diagnosis of self-blast defects of glass insulators based on multi-feature fusion,” Electric power, vol. 50, pp. 52–58, 2017.
[12] J. Yin, Y. Lu, Z. Gong, Y. Jiang, and J. Yao, “Edge detection of high-voltage porcelain insulators in infrared image using dual parity morphological gradients,” IEEE Access, vol. 7, no. 7, Article ID 2900658, pp. 32728–32734, 2019.
[13] S. Anjum, S. Jayaram, A. El-Hag, and A. N. Jahromi, “Detection and classification of defects in ceramic insulators using RF antenna,” IEEE Transactions on Dielectrics and Electrical Insulation, vol. 24, no. 1, Article ID 005867, pp. 183–190, 2017.
[14] H. Cheng, Y. Zhai, R. Chen, D. Wang, Z. Dong, and Y. Wang, “Self-Shattering defect detection of glass insulators based on spatial features,” Energies, vol. 12, no. 3, p. 543, 2019.
[15] Y. Yang, L. Wang, Y. Wang, and X. Mei, “Insulator self-shattering detection: a deep convolutional neural network approach,” Multimedia Tools and Applications, vol. 78, no. 8, pp. 10097–10112, 2019.
[16] T. Lin and X. Liu, “An intelligent identification system for insulator string defects based on dimension correction and optimized faster R-CNN,” Electr. Eng., vol. 6, pp. 1–9, 2020.
[17] C. Ge, J. Wang, J. Wang, Q. Qi, H. Sun, and J. Liao, “Towards automatic visual inspection: a weakly supervised learning method for industrial applicable object detection,” Computers in Industry, vol. 121, Article ID 103232, 2020.
[18] Q. Wen, Z. Luo, R. Chen, Y. Yang, and G. Li, “Deep learning approaches on defect detection in high resolution aerial images of insulators,” Sensors, vol. 21, no. 4, p. 1033, 2021.
[19] Y. Wang and Z. Li, “Yang X.; Luo N.; zhao Y.; zhou Y.; zhou G. Insulator defect identification based on faster R-CNN,” in Proceedings of the 2020 International Conference on Computer, Information and Telecommunication Systems (CITS), pp. 5–7, Hangzhou, China, October 2020.
[20] X. Liu, H. Jiang, J. Chen, S. Zhang, and X. Miao, “Insulator detection in aerial images based on faster regions with convolutional neural network,” in Proceedings of the 2018 IEEE 14th International Conference on Control and Automation (ICCA), pp. 12–15, IEEE, AK, USA, June 2018.

[21] Z. Zhao, Z. Chen, L. Zhang, Y. Qi, Y. Kong, and K. Zhang, “Insulator detection method in inspection image based on improved faster R-CNN,” Energies, vol. 12, no. 7, p. 1204, 2019.

[22] L. Huang, Z. Chao, and M. Wang, “Detection and evaluation method of transmission line defects based on deep learning,” IEEE Access, vol. 8, pp. 38448–38458, 2020.

[23] P. Fan, H. Shen, C. Zhao et al., “Defect identification detection research for insulator of transmission lines based on deep learning,” Journal of Physics: Conf. Ser., vol. 1828, no. 1, Article ID 012019, 2021.

[24] M. Tomaszewski, P. Michalski, and J. Osuchowski, “Evaluation of power insulator detection efficiency with the use of limited training dataset,” Applied Sciences, vol. 10, no. 6, p. 2104, 2020.

[25] Y. Yang, Y. Wang, and H. Jiao, “Insulator identification and self-shattering detection based on mask region with convolutional neural network,” Journal of Electronic Imaging, vol. 28, no. 5, p. 1, 2019.

[26] X. Tao, D. Zhang, Z. Wang, X. Liu, H. Zhang, and D. Xu, “Detection of power line insulator defects using aerial images analyzed with convolutional neural networks,” IEEE Trans. Syst. Man Cybern. Syst., vol. 50, no. 4, Article ID 2871750, pp. 1486–1498, 2020.

[27] F. Gao, J. Wang, Z. Kong et al., “Identification of insulator explosion based on deep learning,” in Proceedings of the 2017 14th international computer conference on wavelet active media technology and information processing ICCWAMTIP, Chengdu, sichuan province, China, December, 2017.

[28] X. Li, H. Su, and G. Liu, “Insulator defect identification based on global detection and local segmentation,” IEEE Access, vol. 8, no. 99, 1 page, 2020.

[29] J. Redmon and A. Farhadi, “YOLO9000: better, faster, stronger,” in Proceedings of the IEEE conference on computer vision and pattern identification, pp. 7236–7271, Honolulu, HI, USA, July 2017.

[30] J. Redmon and A. Farhadi, “Yolov3: an incremental improvement,” 2018, https://arxiv.org/abs/1804.02773, 1804.02773.

[31] K. He, X. Zhang, and S. Ren, “Identity mappings in deep residual networks,” in Proceedings of the 2016 European Conference on Computer Vision, pp. 338–355, Heidelberg, Germany, October 2016.

[32] J. Han, Z. Yang, Q. Zhang et al., “A method of insulator faults detection in aerial images for high-voltage transmission lines inspection,” Applied Sciences, vol. 9, no. 10, p. 2019, 2019.

[33] C. Liu, Y. Wu, and Z. Sun, “Insulator faults detection in aerial images with single shot multibox detector,” IEEE Access, vol. 8, Article ID 2891123, pp. 9945–9956, 2019.

[34] D. Sadykova, D. Pernebaveya, M. Bagheri, and A. James, “IN-YOLO: real-time detection of outdoor high voltage insulators using UAV imaging,” IEEE Transactions on Power Delivery, vol. 35, no. 3, Article ID 2944741, pp. 1599–1601, 2020.

[35] M. W. Adou, H. Xu, and G. Chen, “Insulator faults detection based on deep learning,” in Proceedings of the International Conference on Anti-counterfeiting, Security, and identification, pp. 25–27, Xiamen, China, October 2019.

[36] C. Xu, B. Bo, Y. Liu, and F. Tao, “Detection method of insulator based on single shot MultiBox detector,” Journal of Physics: Conf. Ser., vol. 1069, Article ID 012183, 2018.

[37] J. Tan, “Automatic insulator detection for power line using aerial images powered by convolutional neural networks,” Journal of Physics: Conf. Ser., no. 4, Article ID 042012, 1748 pages, 2021.

[38] C. Liu, Y. Wu, J. Liu, and Z. Sun, “Improved YOLOv3 network for insulator detection in aerial images with diverse background interference,” Electronics, vol. 10, no. 7, p. 771, 2021.

[39] J. Han, Z. Yang, Q. Zhang et al., “A method of insulator faults detection in aerial images for high-voltage transmission lines inspection,” Applied Sciences, vol. 9, no. 10, p. 2019, 2019.

[40] C. Liu, Y. Wu, J. Liu, Z. Sun, and H. Xu, “Insulator faults detection in aerial images from high-voltage transmission lines based on deep learning model,” Applied Sciences, vol. 11, no. 10, p. 4647, 2021.

[41] B. Krawczyk, “Learning from imbalanced data: open challenges and future directions,” Prog. Artif. Intell., vol. 5, no. 4, pp. 221–232, 2016.

[42] K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks,” in Proceedings of the 14th European Conference on Computer Vision, pp. 630–6453–7, AZ, USA, November 2014.

[43] G. Huang, Z. Liu, L. Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” 2016, https://arxiv.org/abs/1608.06993.

[44] C. Wang, H. Liao, Y. Wu, P. Chen, and I. Yeh, “CSPNet: a new backbone that can enhance learning capability of CNN,” in Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern identification Workshops (CVPRW), IEEE, WA, USA, June 2020.

[45] K. He, X. Zhang, S. Ren, and J. Sun, “Spatial pyramid pooling in deep convolutional networks for visual recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 9, Article ID 2389824, pp. 1904–1916, 2015.

[46] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, Path Aggregation Network for Instance Segmentation, IEEE, in Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern identification (CVPR), pp. 18–22, June 2018.

[47] YOLOv4, “Project of the dark-net framework,” 2020, https://github.com/AlexeyAB/darknet.

[48] J. Liu, C. Liu, Y. Wu, H. Xu, and Z. Sun, “An improved method based on deep learning for insulator fault detection in diverse aerial images,” Energies, vol. 14, p. 4365, 2021.