

Received 4 March 2025, accepted 3 April 2025, date of publication 10 April 2025, date of current version 5 May 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3559657

## RESEARCH ARTICLE

# Defects Localization and Classification Method of Power Transmission Line Insulators Aerial Images Based on YOLOv5 EfficientNet and SVM

LIN LI<sup>1</sup>, QIAOLING YIN<sup>2</sup>, XIAOFENG WANG<sup>3</sup>, AND HANG WANG<sup>1</sup>, (Member, IEEE)

<sup>1</sup>Hubei Engineering Research Center for Safety Monitoring of New Energy and Power Grid Equipment, Hubei University of Technology, Wuhan 430068, China

<sup>2</sup>State Grid Hefei Electric Power Supply Company, Anhui Electric Power Company Ltd., Hefei 230061, China

<sup>3</sup>CSG Guangdong Chaozhou Power Supply Bureau, Chaozhou 519000, China

Corresponding author: Hang Wang (wanghang@whu.edu.cn)

**ABSTRACT** Accurate localization and classification of defects in insulator images captured by UAVs (Unmanned Aerial Vehicles) are critical for the maintenance of high-voltage power transmission grids. However, existing methods face challenges in achieving satisfactory accuracy, particularly in complex environments. Moreover, previous approaches primarily focus on detecting a single type of insulator defect without further categorizing them. To address these limitations, we propose a novel deep-learning framework that integrates YOLOv5, EfficientNet and Support Vector Machines (SVM) to improve both the localization and classification of insulator defects in aerial images. The framework specifically targets overcoming the challenges associated with low signal-to-noise ratios in defect detection. The proposed approach divides the task into two primary modules: 1) YOLOv5-based object detection for accurate defect localization, and 2) defect classification using EfficientNet and SVM. Specifically, YOLOv5 enhances feature extraction to improve the signal-to-noise ratio, while EfficientNet and SVM enable precise defect classification. The effectiveness of the proposed method is validated through experiments conducted on a range of insulator image datasets collected from the Guangdong power grid. Comparative analysis of experimental results shows that the proposed method outperforms existing approaches in terms of both accuracy and practical applicability, achieving a top-1 accuracy of 98.22%, an F1-score of 98.53%, and an accuracy improvement of 3.89% over conventional backbone-based networks.

**INDEX TERMS** Insulator images, defects classification, deep learning, YOLOv5, EfficientNet, SVM.

## I. INTRODUCTION

High-voltage electrical power transmission systems have been rapidly growing in the last decades. Insulators between the lines and towers provide mechanical support and electrical insulation in power systems. However, insulators are vulnerable to environmental factors, leading to wear and damage, particularly under extreme weather conditions such as corrosion from lightning strikes and rainwater. Such damage can have a significant impact on the safety and operational stability of the power system [1], [2], [3]. To ensure the reliable operation of overhead transmission lines, regular inspections of insulators are therefore

essential. Manual inspections, helicopter-based surveys and drone-assisted inspections are currently the main inspection methods for these lines [4]. Manual inspection of overhead transmission lines often encounters high complexity and low efficiency due to the limitations of line distribution and complex environmental conditions. In contrast, while helicopter inspection can cover a larger area, its high cost and relatively high risk present significant limitations in practical applications. Methods for automatically detecting transmission line faults using aerial images or video captured by drones are becoming more common with the rapid development of drone technology. Given the substantial number of inspection images obtained from drones, the existing approaches for identifying insulator defects can generally be divided into two main types: the first employs

The associate editor coordinating the review of this manuscript and approving it for publication was Jon Atli Benediktsson<sup>1</sup>.

conventional image processing methods, whereas the second leverages deep learning algorithms for detecting specific targets.

Threshold segmentation [5], morphological operations [6], and edge detection [7] have historically served as foundational techniques in this domain. However, with the rapid advancements in deep learning methodologies, these conventional approaches are progressively being replaced by deep learning-based techniques, which are emerging as the dominant methodology in the field [8]. By leveraging neural networks and large-scale training datasets, deep learning methods have been shown to be highly capable of detecting defects with both speed and accuracy, while improving robustness and generalizability.

This section reviews object detection methods based on deep learning, covering both single-stage detectors [9] and two-stage detectors [10]. Single-stage detectors, such as SSD [11] and YOLO [12], [13], [14], are widely applied due to their simplicity and high efficiency. In recent years, extensive research has focused on enhancing the performance of these classic networks. For example, [15] incorporated EIoU and a novel loss function into YOLOv3, which significantly improved both the alignment accuracy of predicted boxes and the model's convergence speed, ultimately enhancing detection precision and efficiency. Another work [17] introduced YOLOv5-tassel, which integrates a bidirectional feature pyramid and SimAM attention module, enhancing small object detection. However, its target application is agricultural crop detection rather than insulator defects. Further improvement was proposed in [18], where the authors developed MFI-YOLO based on YOLOv8, integrating a C2F network and ResPANet to improve multi-object detection under complex backgrounds. Despite its enhanced performance, the reported accuracy remained at 93.9%, leaving room for improvement. In [19], a lightweight detection framework was developed by combining SSD with MnasNet to detect insulators and spacers in real-time UAV inspections. This approach achieved an accuracy of 93.8%, demonstrating good computational efficiency. Similarly, [20] proposed an improved defect detection method based on CenterNet, which simplifies the backbone network and incorporates a super-resolution module to enhance image quality and detection precision. However, this work only focuses on detecting a single defect type. To further enhance detection in complex aerial images, [21] developed a detection framework based on CSP-ResNeSt and Bi-SimAM-FPN, which significantly improves the detection of small insulator defects. However, this method still limits its focus to a single defect type. Finally, [22] introduced an enhanced SSD model for detecting insulators in infrared images of substations, incorporating multi-scale feature fusion and transfer learning. Nevertheless, this method still struggles with missed and false detections, especially for small or occluded defects.

With the development of two-stage detectors, considerable efforts have been made to enhance their detection performance. For instance, the authors in [23] proposed a defect detection system combining Faster R-CNN with a deep multi-task network for insulator surface defect identification, although false positives and false negatives remain problematic. Another study [24] employed Faster R-CNN for insulator detection in complex backgrounds, integrating image cropping and adaptive HSV thresholding for preprocessing. Despite achieving good accuracy, the method requires high-quality images and suffers from slow inference speed. Other studies explored different strategies to improve performance. For example, [25] applied Faster R-CNN for fault detection in high-voltage transmission lines but fell short of real-time requirements. To improve adaptability, [26] combined Mask R-CNN, image processing, feature fusion, and K-means clustering for defect detection in railway insulators, although its applicability to other insulator types remains limited. A more efficient solution was introduced in [27], where Faster R-CNN with EfficientNet replaced the conventional ResNet backbone, significantly reducing parameters and enhancing detection accuracy and speed. Transfer learning was further employed to improve the detection of small defects. In addition, [28] proposed Cascade R-CNN to enhance object detection and instance segmentation by progressively increasing IoU thresholds to improve sample quality and reduce overfitting.

In summary, the accurate identification of insulator defects in complex backgrounds still faces two major technical challenges, which this study aims to overcome through a novel detection and classification framework:

#### (1) Multi-Category Defect Identification Capability

Most existing methods only detect a single type of insulator defect, such as self-explosion or breakage, without further classifying them. This narrow focus limits their practical applicability in real inspections, where self-explosion, breakage, and flashover often coexist, requiring comprehensive multi-class defect identification.

In contrast, the proposed method first employs YOLOv5 to detect insulators and defect regions, then utilizes EfficientNet+SVM to achieve fine-grained classification of the detected defect images. Meanwhile, a comprehensive insulator defect dataset is constructed, covering multiple insulator types (glass, ceramic, composite) and multiple defect categories (self-explosion, breakage, flashover). This two-stage strategy not only enhances detection accuracy but also enables precise multi-class defect identification, significantly improving the model's adaptability to diverse inspection scenarios.

#### (2) Low Signal-to-Noise Ratio Due to Small Target

In UAV inspection images, insulators occupy only a small portion of the image, leading to a low signal-to-noise ratio (SNR), which poses challenges for both detection and classification in conventional approaches.

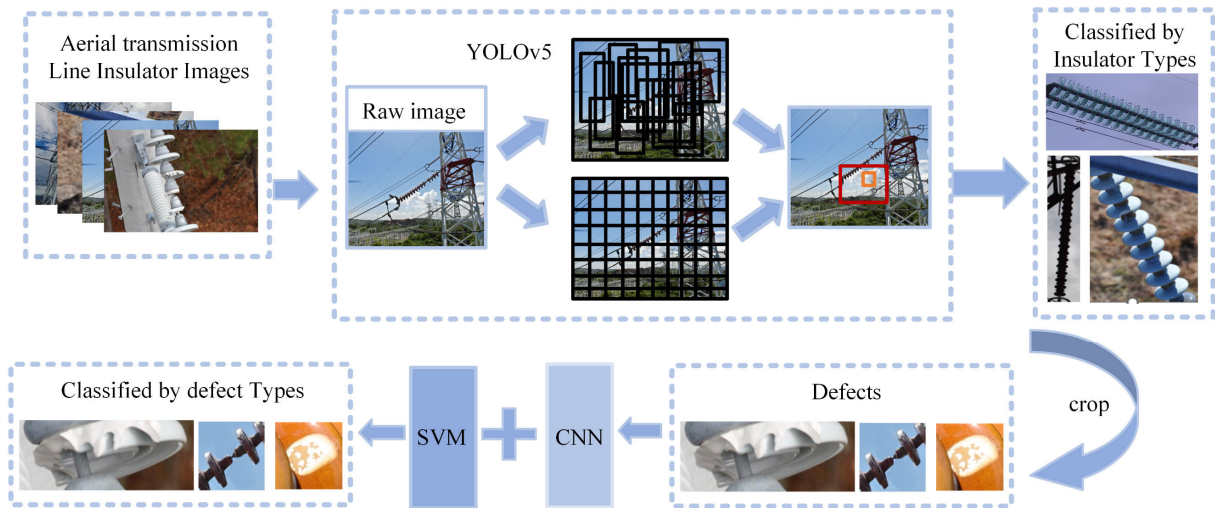


FIGURE 1. Overall flow chart of the proposed method.

To address these challenges, this paper proposes a novel detection and classification framework for insulator defects in UAV images, which decomposes the insulator detection task into two key computer vision problems: object detection and image classification [29], [30], [31], [32], [33]. First, YOLOv5 performs accurate detection and localization of insulators and their defects, leveraging its multi-scale feature fusion capability to enhance small object detection. The detected defect regions are then cropped to remove irrelevant background, significantly improving the SNR of the defect images. These high-SNR defect images are subsequently classified by EfficientNet+SVM, enabling high-precision classification across multiple defect types. This hybrid design effectively leverages the complementary advantages of deep learning and traditional machine learning, enabling high-accuracy defect detection and classification.

## II. METHODOLOGY FOR LOCATING AND CLASSIFYING INSULATORS AND INSULATOR DEFECTS

Regarding the localization and classification methodology for insulators and insulator defects, several aspects need consideration:

### 1) Object detection method for insulators and defects:

Choosing an efficient object detection method is essential for identifying and locating insulators and their defects. While numerous object detection algorithms are available, they lack universality across different datasets. Therefore, an appropriate object detection algorithm must be selected to establish the foundation for subsequent classification tasks on insulator strings and insulator defects.

### 2) Improvement of Signal-to-Noise ratio:

Cropping the portion that contains the insulator strings enhances the signal-to-noise ratio. This transforms the problem into a positional issue for insulator strings and defects. Given YOLO's rapid and efficient characteristics in

object detection and localization, the YOLO series algorithm is selected to address the positioning of insulators and defects.

### 3) Selection of classification method:

When faced with image classification tasks, one can opt for machine learning-based classification methods or convolutional neural networks (CNNs). Accurately classifying images of insulator defects is a critical challenge. In this study, the Softmax layer responsible for classification in the CNN network is replaced with SVM to achieve precise classification of insulator defects.

Based on the above considerations, this study proposes a localization and classification framework for insulators and their defects in overhead transmission lines. The structure of the proposed framework is illustrated in Figure 1. The framework consists of two main modules: 1) an object detection module based on YOLOv5, responsible for accurately locating insulator defects, and 2) an image classification module using CNNs combined with SVM for defect type recognition. Cropped defect images are fed into the CNN+SVM module to generate the final classification output.

## III. INSULATOR STRING AND DEFECT DETECTION

### A. YOLOV5

This study adopts the YOLOv5 single-stage object detection network to identify and locate insulation defects. By leveraging YOLOv5 for object detection, it not only enables efficient identification of insulator defects but also lays a solid foundation for subsequent defect classification. The YOLOv5 network structure is illustrated in Fig. 2, is comprised of four key components: the input stage, the backbone network, the neck network, and the output stage. At the input stage, YOLOv5s incorporates preprocessing techniques such as adaptive image padding and Mosaic data augmentation. Mosaic data augmentation involves combining four randomly selected images through geometric transformations, thereby

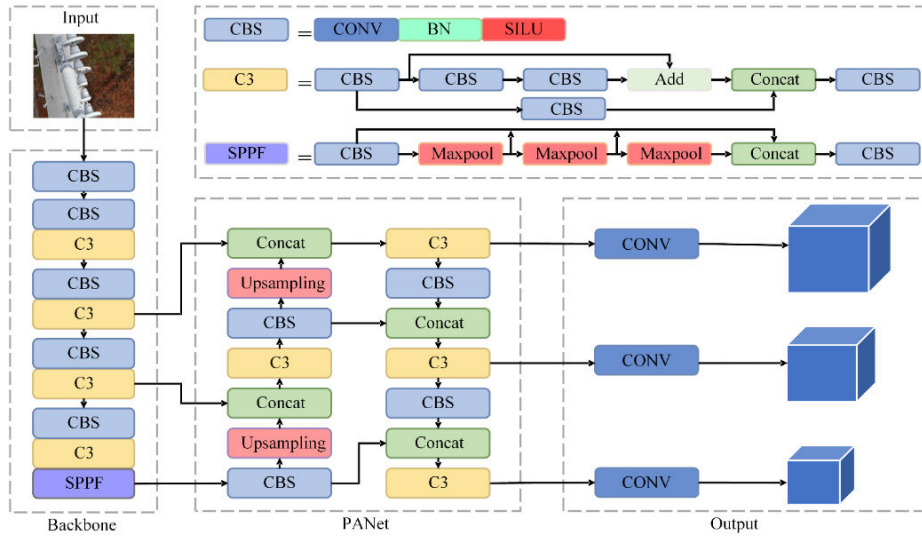


FIGURE 2. Architecture of YOLOv5.

increasing image complexity. This process enhances the model's ability to detect objects in challenging and dynamic environments.

The backbone network of YOLOv5s, CSPDarknet, consists of three critical modules: CBS, C3, and SPPF, all of which are responsible for the extraction of feature representations. The CBS module integrates convolutional operations, normalization, and activation functions. The C3 module, a central element in feature extraction, adopts the CSPNet architecture, where the input feature map is divided into two separate paths, each undergoing convolutional operations. The resulting feature maps from both paths are then concatenated and processed through a subsequent CBS module to produce the final output feature map. The SPPF module improves upon the SPP module by applying three successive max-pooling operations on the convolutional feature map, followed by concatenation of the pooled results. This strategy increases the receptive field, thereby enhancing the model's ability to detect objects at varying scales.

The neck network in YOLOv5s employs a Path Aggregation Network (PANet) structure, which extracts feature maps from the P3 to P5 layers of the backbone network. PANet employs both upsampling and downsampling operations to facilitate the fusion of features across different scales. This feature fusion mechanism enables the model to more effectively combine shallow image features with deep semantic information, thereby improving detection accuracy for objects of diverse sizes and shapes.

At the output stage, YOLOv5s utilizes a convolutional block to perform classification and regression predictions on the multi-scale feature maps. Subsequently, the NMS algorithm is applied to eliminate prediction boxes that fail to meet the CIOU threshold, retaining only the boxes that satisfy the conditions as the final detection results. The detailed

definition of the CIOU loss function is provided below.

$$CIOU = IoU - \frac{\rho^2(b, b^{gt})}{c^2} - \alpha v \quad (1)$$

$$v = \frac{4}{\pi^2} (\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h})^2 \quad (2)$$

$$\alpha = \begin{cases} 0, & IoU < 0.5, \\ \frac{v}{(1 - IoU) + v}, & IoU \geq 0.5. \end{cases} \quad (3)$$

$$\mathcal{L}_{CIOU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (4)$$

The CIOU function, as shown in equation (1), calculates the loss based on several components. Specifically,  $b$  and  $b^{gt}$  represent the center coordinates of the predicted and ground truth bounding boxes, respectively.  $c$  denotes the diagonal length of the smallest enclosing rectangle that fully contains both the predicted and ground truth boxes.  $IoU$  measures the overlap between these two boxes and is used to assess the accuracy of the prediction. The term  $\rho^2$  calculates the Euclidean distance between the centers of the predicted and ground truth boxes, indicating their spatial alignment. Additionally,  $v$ , as shown in equation (2), is a correction factor that adjusts the loss function by considering the shape and orientation of the bounding boxes. Lastly,  $\alpha$  defined in equation (3), is a balancing factor that helps adjust the relative importance of the various components in the loss function during model training. These elements work together to provide a comprehensive assessment of the model's bounding box predictions and improve the accuracy of detection. The final CIOU loss function is given by equation (4).

## B. EVALUATION INDICATORS

Average Precision (AP) serves as a crucial metric for evaluating the effectiveness of target detection models. Precision



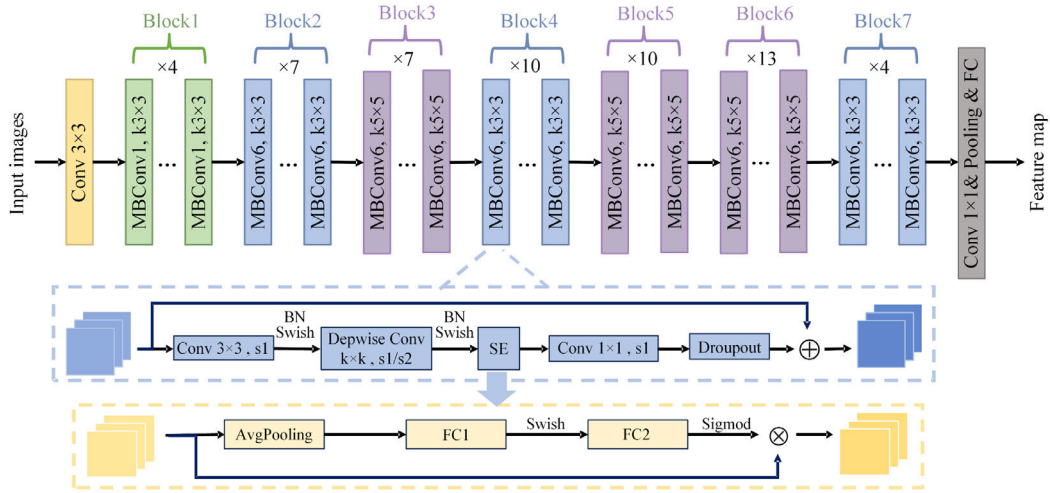


FIGURE 3. Architecture of EfficientNet-B7.

refers to the ratio of defective insulators predicted by the model that are accurately recognized as such, calculated via equation (5). In contrast, recall measures the percentage of actual defective insulator samples that the model manages to identify, as determined by equation (6). When recall is charted along the horizontal axis and the detection rate along the vertical axis, the resulting graph is referred to as the precision-recall (PR) curve. The area contained within this curve, framed by both the x-axis and y-axis, signifies the AP value for identifying insulator defects, which can be computed using equation (7).

$$P = \frac{TP}{TP + FP} \quad (5)$$

$$R = \frac{TP}{TP + FN} \quad (6)$$

$$AP = \int_0^1 P(R) dR \quad (7)$$

#### IV. IMPLEMENTING INSULATOR DEFECT CLASSIFICATION WITH CNNs+SVM

CNNs represent a specific type of artificial neural network architecture that is inspired by the visual cortex of cats. Researchers have developed this network model by mimicking the neural connections observed in the visual system of cats. The architecture is well-suited for simulating the neurons found in the human brain, and it is frequently employed in various image classification tasks. In comparison, SVM [34] operates by effectively separating data points from distinct categories through the identification of an optimal hyperplane. The underlying concept of SVM is to maximize the margin between the support vectors and the hyperplane, offering significant advantages when dealing with both linear and non-linear datasets. In this study, we combine CNNs with the traditional machine learning classification method, SVM, to further enhance the classification of insulator defects.

#### A. EFFICIENTNET

EfficientNet is a network series released by Google in May 2019, designed to improve model accuracy while maintaining a relatively low number of model parameters.

The EfficientNet team found that network depth, network width, and feature resolution all impact the final performance of the network. Theoretically, increasing network depth, enlarging network width, or expanding input image resolution can infinitely improve the performance of the network model. However, as the network depth, width, and feature resolution increase, the performance improvement becomes slower, and it introduces certain challenges to the training process. EfficientNet proposes constraints on three dimensions of network models: depth, channel width of feature maps, and spatial resolution of feature maps, as represented by the following formula:

$$d = \alpha^\phi \quad (8)$$

$$w = \beta^\phi \quad (9)$$

$$r = \gamma^\phi \quad (10)$$

$$\text{s.t. } \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 (\alpha \geq 1, \beta \geq 1, \gamma \geq 1) \quad (11)$$

In this context,  $\Phi$  is a hyperparameter, while  $d$ ,  $w$ , and  $r$  represent depth, width, and resolution, as defined in Equations (8), (9), and (10).  $\alpha$ ,  $\beta$  and  $\gamma$  are employed to govern the magnitudes of depth, width, and resolution, subject to the constraints outlined in Equation (11). EfficientNet employs precision and floating-point operation complexity as search objectives, continuously exploring values for  $d$ ,  $w$ , and  $r$  to determine the network model. Consequently, it introduces eight baseline networks, namely EfficientNet-b0 through EfficientNet-b7.

EfficientNet-b7 stands out among the eight baseline models due to its remarkable capacity and performance. In tasks related to image analysis, this model excels in feature extraction, adeptly identifying intricate details in complex



FIGURE 4. Data augmentation display diagram.

images, which significantly enhances its recognition accuracy. Consequently, this study opts to utilize EfficientNet-b7 as the feature extraction network.

As depicted in Figure 3, the network architecture is segmented into nine separate stages. The first stage includes a typical convolution layer, utilizing a  $3 \times 3$  kernel with a stride length of 2, primarily aimed at reducing the input size while extracting initial features. From the second to the eighth stages, the MBConv block is repeatedly stacked to facilitate multi-level feature extraction. The final stage, designated as Stage 9, contains a traditional  $1 \times 1$  convolutional layer for feature refinement, which is followed by average pooling and a fully connected layer to deliver the final output of the network.

The MBConv block comprises several core components. It begins with a  $1 \times 1$  pointwise convolution to increase the number of feature dimensions, along with batch normalization (BN) and the Swish activation function to introduce non-linearity. Next, a depthwise convolution is performed using a kernel size of  $k \times k$ , where the value of  $k$  can either be 3 or 5, as outlined in Figure 3. The Squeeze-and-Excitation (SE) mechanism is then applied to recalibrate the feature map channels, enhancing the network's capacity to focus on significant features. Afterward, a second pointwise convolution is employed to reduce the dimensionality, followed by a Dropout layer to minimize the risk of overfitting.

The SE module, a crucial component of MBConv, starts with a global average pooling layer that aggregates spatial information. This is succeeded by two fully connected layers designed to refine the channel-wise attention. The first fully connected layer compresses the input channels to a quarter of their original size, leveraging the Swish function to improve the network's non-linear representation. The second fully connected layer restores the channel count to match the output of the depthwise convolution and uses a Sigmoid activation function to produce scaling coefficients for each channel, thereby adjusting their importance dynamically.

## B. SUPPORT VECTOR MACHINE

This paper combines EfficientNet-B7 [35] with a Linear SVM for image classification. In comparison to other classifiers such as neural networks, classification tree model, and kNN, SVM demonstrates superior performance on small datasets and nonlinear high-dimensional features. Subsequently, the final softmax classifier is replaced with a one-vs-all SVM using an RBF kernel. The optimization process incorporates a constant  $C$  that regulates the trade-off between maximizing the margin and minimizing classification errors by applying penalties to misclassified samples. Equation (14) formally defines the corresponding nonlinear decision boundary.

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^j \sum_{j=1}^l y_i y_j \alpha_i \alpha_j K(x, x_j) - \sum_{j=1}^l \alpha_j \quad (12)$$

$$\text{s.t.} \sum_{i=1}^l y_i \alpha_i = 0, 0 \leq \alpha_i \leq C, i = 1, \dots, l \quad (13)$$

$$f(x) = \text{sign}\left(\sum_{i=1}^n \alpha_i y_i K(x, x_i) + b\right) \quad (14)$$

In this context,  $\alpha$  denotes the Lagrange multiplier,  $y$  signifies the corresponding label, and  $x_k$  indicates a support vector mapped to a feature with 2560 dimensions. The kernel function  $K(x, x_k)$  is defined in Equation (15), where  $\sigma$  adjusts the RBF kernel's width, controlling sensitivity to input variations and data fit.

$$K(x, x_i) = \exp\left(\frac{-\|x - x_i\|^2}{2\sigma^2}\right) \quad (15)$$

## V. EXPERIMENT

In this section, we begin by introducing the dataset and outlining the experimental environment settings. Subsequently, we provide a detailed account of the training and testing processes for the two network architectures. An evaluation of the accuracy of this method is conducted, comparing it

with other approaches. The final results of the proposed method are presented in the form of images, demonstrating the advantages of combining the two models.

### A. DATASETS

The insulator images used in this study are sourced from two main origins: the first consists of drone inspection photos of transmission lines provided by the China Southern Power Grid Company, while the second comprises publicly available datasets from the Internet. The drone inspection photos primarily feature glass insulators and composite insulators. Employing a method of image data cleaning, high-quality images of 1122 transmission line insulators were obtained by removing similar images of insulators and images of transmission lines without insulators. Subsequently, considering real-world insulator defect scenarios, 133 randomly selected images containing complete strings of glass insulators were processed, resulting in 133 images containing self-explosion defects of glass insulators. The remaining 989 images represent transmission line insulator images without defects. The publicly available dataset includes 1631 images of ceramic insulators, primarily featuring defects such as broken damage and flashover damage. Data augmentation [36] techniques were applied to the insulator dataset to enhance its diversity and prevent overfitting during the training phase. Specifically, the original image is processed through random angle rotation, cutout, blurring, brightness modification, and Gaussian noise addition. The specific details are illustrated in Figure 4.

The original dataset comprises 2,724 images collected from two distinct sources. Following the application of data augmentation techniques, the total number of images was increased to 11,565. The augmented dataset was subsequently partitioned into training, validation, and test sets in a 6:2:2 ratio. As a result, the training set includes 6,939 insulator images, while both the validation and test sets contain 2,313 insulator images each.

### B. EXPERIMENTAL SETUP

The experiment utilized the PyTorch framework. The deep learning model was trained using the following hyperparameter settings. The initial learning rate was set to 0.01, and the training process was conducted over 400 epochs. The input image size for the network was standardized to  $640 \times 640$  pixels. A batch size of 16 was employed to optimize the training efficiency. To prevent overfitting, weight decay was set to 0.0005, and the momentum parameter was fixed at 0.937 to accelerate convergence and improve stability during the optimization process.

The experimental setup was conducted on a workstation running Windows 10 as the operating system. The model training and evaluation were accelerated using an NVIDIA GeForce GTX 2080 Ti GPU. The deep learning framework was configured with CUDA 10.1 for GPU-accelerated computations, and OpenCV 3.4.6 was employed for image processing tasks.

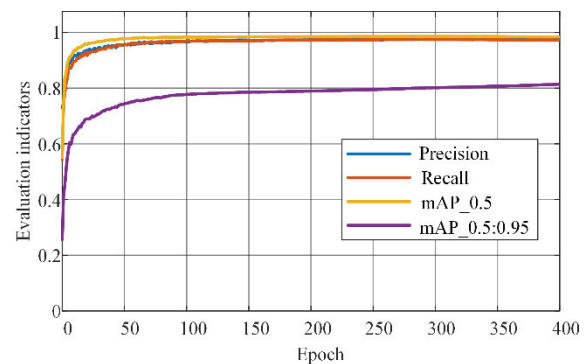


FIGURE 5. YOLOv5 object detection training results.

### C. INSULATOR TARGET DETECTION RESULTS

In this study, YOLOv5 is utilized for the positioning of insulator strings, and the training results are presented in Figure 5. The values of various evaluation metrics in the training results are quite high. Specifically, the Precision value is 0.978, the Recall value is 0.972, the mAP50 value is 0.984, and the mAP95 value is 0.813.

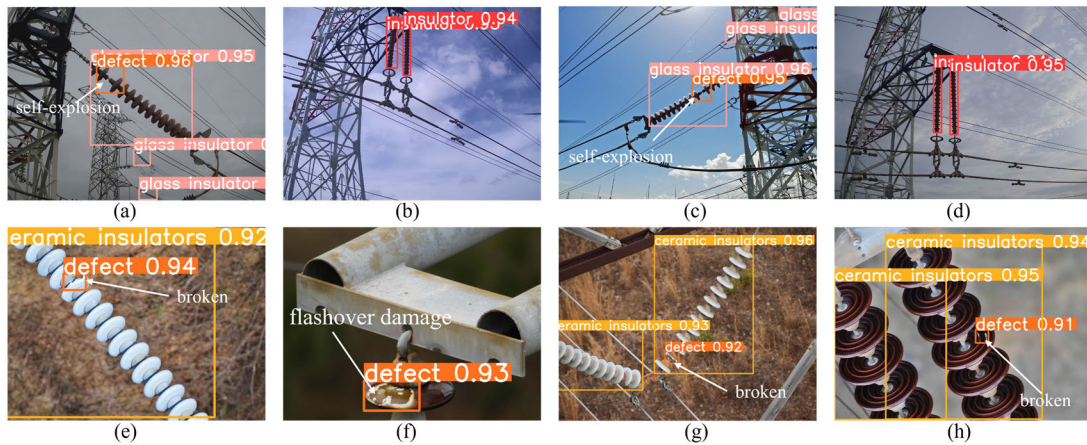
Figure 6 illustrates the localization results for several types of insulators, alongside the precise detection of insulator defects. Each target is highlighted with a rectangular bounding box, annotated with a confidence score, effectively demonstrating the model's detection performance. As depicted in Figure 6, this localization method demonstrates considerable effectiveness in detecting insulator strings and their defects in complex scenes.

### D. INSULATOR DEFECT CLASSIFICATION

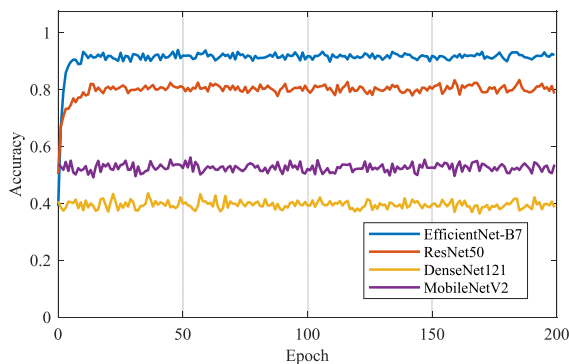
The cropped insulator defect images obtained from YOLOv5 were used to construct a dataset for the insulator defect classification task, totaling 20,265 images. These images were categorized into three types: insulator breakage (7,398 images), glass-insulator self-explosion (987 images), and flashover damage (11,880 images). The dataset was then divided into training, validation, and test sets using a 6:2:2 ratio. As a result, the training set contains 12,159 images, while both the validation and test sets each consist of 4,053 images. Using the insulator defect images identified by YOLOv5 as input for the CNN+SVM classification module, figure 7 below presents the results of training several different CNN networks on the insulator defect dataset. Specifically, the average accuracy for EfficientNet-B7 is 0.921, the average accuracy for ResNet50 is 0.785, the average accuracy for DenseNet121 is 0.405, and the average accuracy for MobileNetV2 is 0.535. The experimental results demonstrate that EfficientNet-B7 outperforms the other networks, achieving the highest average accuracy among the tested models. Based on this superior performance, EfficientNet-B7 was selected as the feature extractor for the insulator defect classification task.

The confusion matrix serves as a visual aid for assessing classification outcomes. It is primarily used to compare





**FIGURE 6.** (a) Detection results of glass insulators and defects; (b) Detection results of insulators; (c) Detection results of glass insulators and defects; (d) Detection results of insulators; (e) Detection results of ceramic insulators and defects; (f) Detection results of ceramic insulators and defects; (g) Detection results of ceramic insulators and defects; (h) Detection results of ceramic insulators and defects.



**FIGURE 7.** Accuracy comparison of EfficientNet-B7, ResNet50, DenseNet121, and MobileNetV2 on the insulator defect dataset, showing training accuracy across epochs.

**TABLE 1.** The structure of the confusion matrix.

Condition	Predicted Condition	
	Prediction positive	Prediction negative
Condition positive	TP(True Positive)	FN(False Negative)
Condition negative	FP(False Positive)	TN(True Negative)

classification outcomes in the assessment of image classification accuracy.

The structure of a confusion matrix is shown in Table 1.

In a confusion matrix, each column shows a predicted class, with the sum of values indicating the total samples assigned to that class by the model. Meanwhile, the rows denote the actual categories of the data, with the sum of each row representing the count of instances for that specific category. The figures within the columns illustrate the count of true instances accurately predicted as belonging to that category.

**TABLE 2.** Ablation study table.

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
EfficientNet-B7 + FC	94.33	94.52	94.66	94.37
ResNet50 + SVM	91.81	94.74	87.59	90.69
EfficientNet-B7 + SVM	<b>98.22</b>	<b>98.78</b>	<b>98.28</b>	<b>98.53</b>
EfficientNet-B7 + RandomForestClassifier	91.26	94.33	86.41	89.79
EfficientNet-B7 + ExtraTreesClassifier	90.99	94.42	83.16	87.63
EfficientNet-B7 + MLPClassifier	97.63	98.27	97.90	98.09
EfficientNet-B7+KNN	95.85	96.74	96.04	96.34

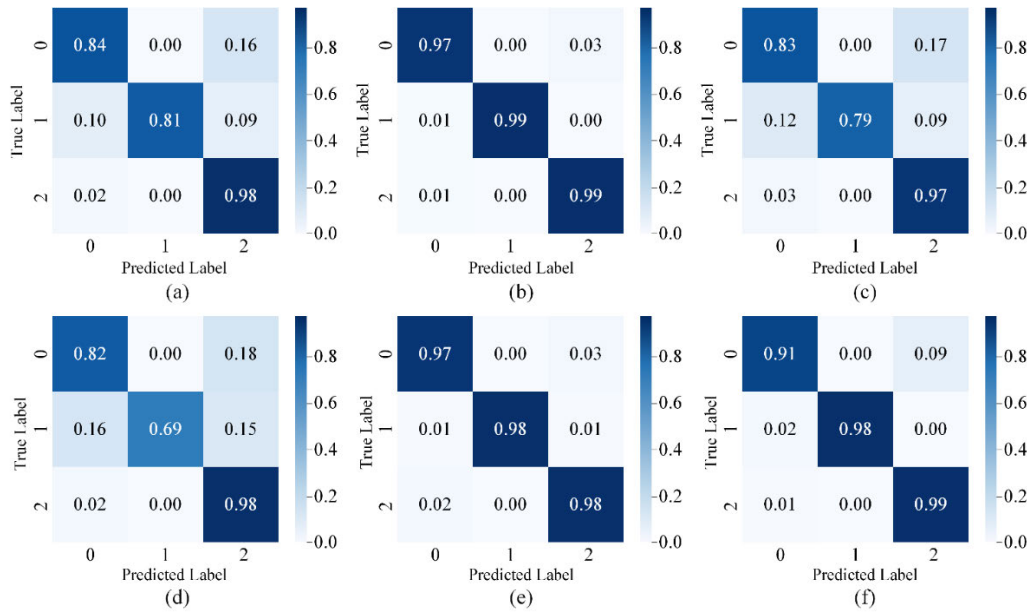
For a classification task involving three categories, the confusion matrix is a  $3 \times 3$  matrix, in which the element in the  $i$  row and the  $j$  column represent the probability of the  $i$  class being classified as the  $j$  class. It is clear that the higher the values on the diagonal, the better the classifier is.

Figure 8 presents the confusion matrices obtained from experiments combining different CNN networks with various traditional machine-learning classification methods. In Figure 8, 0 represents broken, 1 represents self-explosion, and 2 represents flashover damage.

To analyze the contribution of different components to the overall performance, we conducted a series of ablation experiments, using EfficientNet-B7 as the baseline model. The classifier module was gradually replaced while keeping all other hyperparameters fixed.

As shown in Table 2, the SVM classifier achieved the highest performance, with an accuracy of 98.22%, representing an improvement of 3.89% compared to the baseline model. This improvement can be attributed to SVM's ability to construct nonlinear hyperplanes in the high-dimensional feature space (2560 dimensions), effectively overcoming the limitations





**FIGURE 8.** (a) Confusion matrix for ResNet50+SVM classification method; (b) Confusion matrix for EfficientNet-B7+SVM classification method; (c) Confusion matrix for EfficientNet-B7+ RandomForestClassifier classification method; (d) Confusion matrix for EfficientNet-B7+ ExtraTreesClassifier classification method; (e) Confusion Matrix for EfficientNet-B7+ MLPClassifier Classification Method; (f) Confusion Matrix for EfficientNet-B7+KNeighborsClassifier classification method.

of the fully connected layer (FC) in EfficientNet-B7, which relies on linear mapping. Notably, tree-based ensemble methods such as Random Forest and ExtraTrees exhibited a significant performance drop under the same feature space, with accuracy decreasing by 3.07%–3.34%. Furthermore, the large discrepancy between precision (94.33%–90.99%) and recall (86.41%–83.16%) revealed a systematic issue of missed defect detection. When replacing the feature extractor with ResNet50, the classification accuracy dropped by 6.41%, further demonstrating that the quality of extracted features sets an upper limit on classification performance. EfficientNet-B7, with its compound scaling strategy, produces more discriminative features than the residual features extracted by ResNet50. Overall, the ablation study confirms that the EfficientNet-B7 + SVM combination consistently outperforms other configurations across accuracy, precision, recall, and F1 score, highlighting its superiority. This indicates that the strong feature extraction capability of EfficientNet-B7 and the effective classification ability of SVM are complementary, jointly contributing to enhanced overall performance.

To demonstrate the advantage of our proposed CNN+SVM approach, we conducted experiments combining different CNN architectures with various traditional classifiers. The resulting confusion matrices were normalized for comparison, as shown in Figure 8. Panel (a) presents the confusion matrix generated by the ResNet50+SVM classification method, with diagonal values of 0.84, 0.81, and 0.98. Panel (b) displays the confusion matrix for the EfficientNet-B7+SVM method, with diagonal values of

0.97, 0.99, and 0.99. Panel (c) illustrates the outcomes from EfficientNet-B7 combined with a RandomForestClassifier [37], yielding diagonal values of 0.83, 0.79, and 0.97. Panel (d) outlines the results using EfficientNet-B7 with an ExtraTreesClassifier [38], where the diagonal values stand at 0.82, 0.69, and 0.98. Panel (e) depicts the confusion matrix from the use of EfficientNet-B7 with an MLPClassifier [39], indicating diagonal values of 0.97, 0.99, and 0.98. Lastly, panel (f) shows the EfficientNet-B7 combined with the KNeighborsClassifier [40] approach, with diagonal values of 0.91, 0.98, and 0.99.

Based on the classification results, this study calculated the precision, accuracy, recall, and F1 score, which are presented in a bar graph, as shown in Figure 9

Following a comprehensive statistical analysis, method b demonstrated a significant advantage in key performance indicators, showcasing a clear superiority over other evaluated methods. A detailed examination reveals that method b achieved an accuracy of 98.223%, recall of 98.288%, precision of 98.775%, and an F1 score of 98.528%, all surpassing the performance of the second-best method, method e, which attained 97.631%, 97.909%, 98.277%, and 98.092% respectively. Specifically, these results indicate that method b outperforms method e by 0.592 percentage points in accuracy, 0.379 percentage points in recall, 0.498 percentage points in precision, and 0.436 percentage points in F1 score. These statistical findings underscore the substantial performance advantage of method b, further solidifying its position as the preferred choice for addressing insulator defect classification tasks.

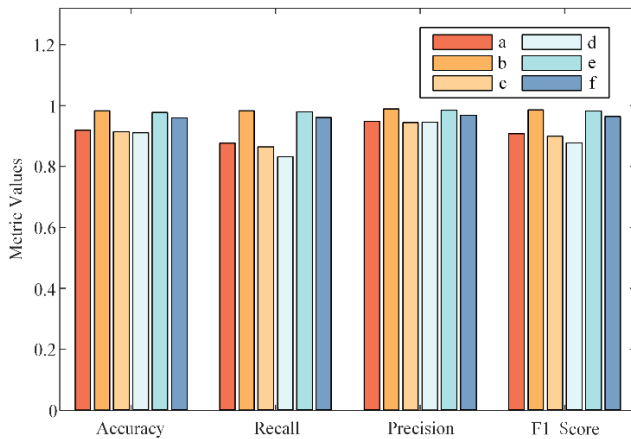


FIGURE 9. Convolutional neural networks training results.

Upon examination, it becomes evident that the EfficientNet-B7+SVM classification method performs exceptionally well across all categories, suggesting it as a highly effective approach. The method we proposed for the insulation defect classification task outperforms all other approaches significantly.

## VI. CONCLUSION

Aerial image processing plays a pivotal role in determining the faults of broken insulators. This paper introduces a novel deep-learning architecture based on aerial images for the localization and classification of insulators and their defects. We transform the problem of identifying insulator defects into a computer vision task of localization and classification. By combining YOLOv5 and EfficientNet-B7, two leading CNN-based models, we devise an innovative approach. The former is responsible for locating insulator strings and defects, significantly enhancing the signal-to-noise ratio. The latter excels in precisely classifying cropped insulator defect images of various sizes. The natural fusion of these two architectures yields distinct advantages in our novel approach. We have selected YOLOv5 as the optimal target detection architecture, achieving the best accuracy performance while ensuring real-time detection: Precision of 0.978, Recall of 0.972, and MAP50 of 0.984. For the insulator defect classification task, we have combined the EfficientNet-B7 architecture with SVM. With its potent feature extraction capabilities, EfficientNet-B7, combined with the remarkable classification abilities of SVM, positions our method as superior to other approaches in insulator defect classification tasks.

Despite the promising results, several limitations remain in our study. First, while the proposed approach demonstrates strong defect detection and classification accuracy, we have not thoroughly analyzed its computational efficiency. Given the complexity of deep learning models, computational cost is a key factor, especially for real-time deployment or large-scale datasets. Future work will focus on improving the

model's computational efficiency and lightweight nature for faster processing and easier integration into practical systems.

Additionally, although our method performs well, the current dataset size for insulator defects is relatively limited, which may affect the comprehensive evaluation of robustness. In future work, we plan to collect a larger and more diverse set of insulator defect images from real-world inspections, further enhancing the system's robustness and improving its ability to handle various defect types and environmental conditions.

Lastly, we aim to extend our approach to other infrastructure monitoring tasks, such as detecting faults in power lines and transformers.

## REFERENCES

- [1] L. She, Y. Fan, J. Wang, L. Cai, J. Xue, and M. Xu, "Insulator surface breakage recognition based on multiscale residual neural network," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021.
- [2] K.-C. Park, Y. Motai, and J. R. Yoon, "Acoustic fault detection technique for high-power insulators," *IEEE Trans. Ind. Electron.*, vol. 64, no. 12, pp. 9699–9708, Dec. 2017.
- [3] Z. Qiu, X. Zhu, C. Liao, D. Shi, and W. Qu, "Detection of transmission line insulator defects based on an improved lightweight YOLOv4 model," *Appl. Sci.*, vol. 12, no. 3, p. 1207, Jan. 2022.
- [4] S. A. Hadiwardoyo, E. Hernández-Orallo, C. T. Calafate, J. C. Cano, and P. Manzoni, "Experimental characterization of UAV-to-car communications," *Comput. Netw.*, vol. 136, pp. 105–118, May 2018.
- [5] Y. Zhang, X. Huang, J. Jia, Y. Zhu, L. Zhao, and X. Zhang, "Detection and condition assessment of icicle bridging for suspension glass insulator by image analysis," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 10, pp. 7458–7471, Oct. 2020.
- [6] S. M. Abid Hasan and K. Ko, "Depth edge detection by image-based smoothing and morphological operations," *J. Comput. Design Eng.*, vol. 3, no. 3, pp. 191–197, Jul. 2016.
- [7] B. Li, L. Qin, F. Zhao, H. Liu, J. Yu, M. He, J. Wang, and K. Liu, "Research on edge detection model of insulators and defects based on improved YOLOv4-tiny," *Machines*, vol. 11, no. 1, p. 122, Jan. 2023.
- [8] X. Wu, D. Sahoo, and S. C. H. Hoi, "Recent advances in deep learning for object detection," *Neurocomputing*, vol. 396, pp. 39–64, Jul. 2020.
- [9] H. Li, Z. Wu, C. Zhu, C. Xiong, R. Socher, and L. S. Davis, "Learning from noisy anchors for one-stage object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 10585–10594.
- [10] L. Mu, X. Xu, Z. Xia, B. Yang, H. Guo, W. Zhou, and C. Zhou, "Autonomous analysis of infrared images for condition diagnosis of HV cable accessories," *Energies*, vol. 14, no. 14, p. 4316, Jul. 2021.
- [11] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis.*, Jan. 2016, pp. 21–37.
- [12] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 779–788.
- [13] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
- [14] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," 2019, *arXiv:1911.09070*.
- [15] Y. Ding and X. Nan, "On edge detection algorithms for water-repellent images of insulators taking into account efficient approaches," *Symmetry*, vol. 15, no. 7, p. 1418, Jul. 2023.
- [16] Z. Yang, Z. Xu, and Y. Wang, "Bidirection-fusion-YOLOv3: An improved method for insulator defect detection using UAV image," *IEEE Trans. Instrum. Meas.*, vol. 71, 2022, Art. no. 3521408.
- [17] W. Liu, K. Quijano, and M. M. Crawford, "YOLOv5-tassel: Detecting tassels in RGB UAV imagery with improved YOLOv5 based on transfer learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 8085–8094, 2022.
- [18] M. He, L. Qin, X. Deng, and K. Liu, "MFI-YOLO: Multi-fault insulator detection based on an improved YOLOv8," *IEEE Trans. Power Del.*, vol. 39, no. 1, pp. 168–179, Feb. 2024.

- [19] X. Liu, Y. Li, F. Shuang, F. Gao, X. Zhou, and X. Chen, "ISSD: Improved SSD for insulator and spacer online detection based on UAV system," *Sensors*, vol. 20, no. 23, p. 6961, Dec. 2020.
- [20] C. Wu, X. Ma, X. Kong, and H. Zhu, "Research on insulator defect detection algorithm of transmission line based on CenterNet," *PLoS ONE*, vol. 16, no. 7, Jul. 2021, Art. no. e0255135.
- [21] K. Hao, G. Chen, L. Zhao, Z. Li, Y. Liu, and C. Wang, "An insulator defect detection model in aerial images based on multiscale feature pyramid network," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022.
- [22] H. Zheng, Y. Sun, X. Liu, C. L. T. Djike, J. Li, Y. Liu, J. Ma, K. Xu, and C. Zhang, "Infrared image detection of substation insulators using an improved fusion single shot multibox detector," *IEEE Trans. Power Del.*, vol. 36, no. 6, pp. 3351–3359, Dec. 2021.
- [23] G. Kang, S. Gao, L. Yu, and D. Zhang, "Deep architecture for high-speed railway insulator surface defect detection: Denoising autoencoder with multitask learning," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 8, pp. 2679–2690, Aug. 2019.
- [24] W. Zhao, M. Xu, X. Cheng, and Z. Zhao, "An insulator in transmission lines recognition and fault detection model based on improved faster RCNN," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–8, 2021.
- [25] X. Lei and Z. Sui, "Intelligent fault detection of high voltage line based on the faster R-CNN," *Measurement*, vol. 138, pp. 379–385, May 2019.
- [26] P. Tan, X. Li, J. Ding, Z. Cui, J. Ma, Y. Sun, B. Huang, and Y. Fang, "Mask R-CNN and multifeature clustering model for catenary insulator recognition and defect detection," *J. Zhejiang Univ.-Sci. A*, vol. 23, pp. 745–756, Jun. 2022.
- [27] Y. Chen, C. Deng, Q. Sun, Z. Wu, L. Zou, G. Zhang, and W. Li, "Lightweight detection methods for insulator self-explosion defects," *Sensors*, vol. 24, no. 1, p. 290, Jan. 2024.
- [28] Z. Cai and N. Vasconcelos, "Cascade R-CNN: High quality object detection and instance segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 5, pp. 1483–1498, May 2021.
- [29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 60, May 2017, pp. 84–90.
- [30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [31] S. Rahman, S. Khan, and F. Porikli, "Zero-shot object detection: Learning to simultaneously recognize and localize novel concepts," *IEEE Trans. Image Process.*, vol. 27, pp. 547–563, 2019.
- [32] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1–9.
- [33] X. Dong, L. Zheng, F. Ma, Y. Yang, and D. Meng, "Few-example object detection with model communication," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 7, pp. 1641–1654, Jul. 2019.
- [34] M. A. Hearst, S. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. Appl.*, vol. 13, no. 4, pp. 18–28, Jul. 1998.
- [35] S.-J. Byeon, J. Park, Y. A. Cho, and B.-J. Cho, "Automated histological classification for digital pathology images of colonoscopy specimen via deep learning," *Sci. Rep.*, vol. 12, no. 1, p. 12804, Jul. 2022.
- [36] M. Kisantal, Z. Wojna, J. Murawski, J. Naruniec, and K. Cho, "Augmentation for small object detection," 2019, *arXiv:1902.07296*.
- [37] Q.-Y. Li, Z.-Y. An, Z.-H. Pan, Z.-Z. Wang, Y.-R. Wang, X.-G. Zhang, and N. Shen, "Severe/critical COVID-19 early warning system based on machine learning algorithms using novel imaging scores," *World J. Clin. Cases*, vol. 11, no. 12, pp. 2716–2728, Apr. 2023.
- [38] E. A. Safavi, "Assessing machine learning techniques in forecasting lumpy skin disease occurrence based on meteorological and geospatial features," *Tropical Animal Health Prod.*, vol. 54, no. 1, pp. 54–55, Feb. 2022.
- [39] G. Minaev, P. Müller, K. Salminen, J. Rantala, V. Surakka, and A. Visa, "A comparison of various algorithms for classification of food scents measured with an ion mobility spectrometry," *Sensors*, vol. 21, no. 2, p. 361, Jan. 2021.
- [40] G. Ren, L. Liu, P. Zhang, Z. Xie, P. Wang, W. Zhang, H. Wang, M. Shen, L. Deng, Y. Tao, X. Li, J. Wang, Y. Wang, and X. Wu, "Machine learning predicts recurrent lumbar disc herniation following percutaneous endoscopic lumbar discectomy," *Global Spine J.*, vol. 14, no. 1, pp. 146–152, Jan. 2024.



**LIN LI** was born in Wuhan, Hubei, China, in 1999. She received the B.Sc. degree in electrical engineering from Hubei University of Technology, in 2021, where she is currently pursuing the M.Sc. degree. Her primary research interests include insulator defect detection and infrared image-based defect detection for high-voltage cable accessories.



**QIAOLING YIN** was born in Anhui, China, in 1992. She received the B.Sc. degree in electrical engineering and automation from Hefei University of Technology, in 2015, and the M.Sc. degree in electrical engineering from Wuhan University, in 2018. Since 2018, she has been an Engineer with State Grid Hefei Electric Power Supply Company. Her main research interest includes electrical insulation.



**XIAOFENG WANG** was born in Shantou, Guangdong, China, in 1983. He received the B.Eng. and M.Eng. degrees in electrical engineering from the School of Electrical Engineering, Wuhan University, in 2007 and 2009, respectively. Since 2009, he has been with CSG Guangdong Chaozhou Power Supply Bureau, where he is currently engaged in research on distribution network lightning protection, intelligent distribution networks, and power grid planning.



**HANG WANG** (Member, IEEE) was born in Shandong, China, in 1989. He received the B.Sc. and M.Sc. degrees in software engineering, and the Ph.D. degree in electrical engineering from Wuhan University, in 2010, 2014, and 2019, respectively. Since then, he has been with Hubei University of Technology as a Lecturer. In 2023, he joined the Institute of Electrical Engineering (IEE), Chinese Academy of Sciences (CAS), as a Senior Visiting Scholar. His research interests include discharge plasma physics, measurement, and applications.

...