



ARTICLE

A Simple and Effective Surface Defect Detection Method of Power Line Insulators for Difficult Small Objects

Xiao Lu^{1,*}, Chengling Jiang¹, Zhoujun Ma¹, Haitao Li² and Yuexin Liu²

¹State Grid Jiangsu Electric Power Co., Ltd., Nanjing, 210024, China

²State Grid Changzhou Power Supply Company, Changzhou, 213003, China

*Corresponding Author: Xiao Lu. Email: lx_sz@js.sgcc.com.cn

Received: 06 November 2023 Accepted: 02 February 2024 Published: 25 April 2024

ABSTRACT

Insulator defect detection plays a vital role in maintaining the secure operation of power systems. To address the issues of the difficulty of detecting small objects and missing objects due to the small scale, variable scale, and fuzzy edge morphology of insulator defects, we construct an insulator dataset with 1600 samples containing flashovers and breakages. Then a simple and effective surface defect detection method of power line insulators for difficult small objects is proposed. Firstly, a high-resolution feature map is introduced and a small object prediction layer is added so that the model can detect tiny objects. Secondly, a simplified adaptive spatial feature fusion (S-ASFF) module is introduced to perform cross-scale spatial fusion to improve adaptability to variable multi-scale features. Finally, we propose an enhanced deformable attention mechanism (EDAM) module. By integrating a gating activation function, the model is further inspired to learn a small number of critical sampling points near reference points. And the module can improve the perception of object morphology. The experimental results indicate that concerning the dataset of flashover and breakage defects, this method improves the performance of YOLOv5, YOLOv7, and YOLOv8. In practical application, it can simply and effectively improve the precision of power line insulator defect detection and reduce missing detection for difficult small objects.

KEYWORDS

Insulator defect detection; small object; power line; deformable attention mechanism

1 Introduction

Insulators, as one of the key equipment, are important for secure power transmission. Its main function is to be used together with the metal fittings to fix the conductors to the tower poles and to keep the insulation between the conductors and the tower poles. However, there are various insulator defects due to harsh environments. The insulation characteristics of the defect insulator gradually deviate from the reliable range until it loses its insulation capacity. Common insulator defects include missing caps, pollution, flashover, breakage, and current leakage. According to research statistics, insulator defects cause more than half and the largest proportion of power accidents [1]. Flashover of insulators causes the second largest number of insulator accidents [2]. Nevertheless, insulator status detection is considered one of the most formidable issues within the realm of power line inspection [3].



To prevent power accidents and ensure the stability of the power supply, power companies must regularly inspect the power system to identify and address any equipment defects. In recent years, the inspection ways have undergone rapid development. Traditional inspection ways include (1) maintenance personnel observation inspection. Manual operation is limited by weather, terrain, and environment, which is inefficient and risky. (2) Using a manned helicopter to shoot videos. The helicopter is flown at a safe distance away from power lines and equipment, and then maintenance personnel take videos of various equipment in the power system for subsequent inspection. But this way is costly and inaccurate. As Unmanned Aerial Vehicle (UAV) technology has been widely used in many fields, such as agriculture and search actions, it has also been promoted in power inspection. UAV inspection offers the benefits of low cost and high efficiency. It can capture all kinds of data efficiently and quickly, such as visible image data. By processing the images, and then using the detection and classification technology based on computer vision, the automatic processing and analysis of images can be achieved so that the intelligent surveillance of the power system can be further realized.

However, after several years of development, the vast amounts of aerial images captured by UAVs have become one of the bottlenecks limiting the intelligent monitoring of power systems. Effectively employing advanced technologies for the intelligent processing and analysis of large-scale images has emerged as a critical concern.

Considerable work has been put into the detection of insulators and their defects to better monitor the power system. Zhang et al. [4] introduced an enhanced YOLOv8s model with multi-scale large kernel attention and lightweight Group Shuffle Convolution (GSCConv) to tackle challenges related to sluggish recognition speed and low accuracy. Zhang et al. [5] proposed a densely connected feature pyramid based on YOLOv3. The method can realize the efficient fusion between the positional information of shallow features and the semantic information of deep features. However, it has a large performance gap in object detection with different scales. Jiang [6] proposed a method based on YOLOv5. They adopted a cascade framework to detect all insulator objects and various types of defects in images using the first-level model and second-level model successively. Although the method can detect defects such as flashovers, breakages, and missing caps, its ability to localize the flashovers and breakages with weak edge features is poor and detection precision is low. Xu [7] proposed a super-resolution generative network. They combined GridMask, random erasure algorithm, and adversarial generative network to expand the small object dataset while realizing the boundary clarity of small objects. Then the precision of small object detection was improved by introducing a Transformer and Swin Transformer to improve YOLOv5. However, the structure of the method is more complicated, which is not favorable for practical application.

To address the aforementioned challenges, this paper establishes a dataset encompassing flashover and breakage defects. Additionally, it presents a method for detecting insulator defects in power lines based on YOLOv5. The key contributions of our research can be outlined as follows:

- We introduce a high-resolution feature map to be fused with feature pyramids to enhance the detailed features of small objects in the neck network. Then the small object prediction layer is added to further improve the detection precision of tiny objects.
- We propose an S-ASFF module to improve adaptability to variable scales. It can enhance scale perception for objects by establishing relationships between dimensions of scale and space to retain key information and learn to lead important scale features into a dominant position.
- We propose an EDAM module to highlight weak edge features of insulator defects. It can further inspire the deformable attention mechanism to learn a small number of critical sampling points near the reference point to extract discriminative features.

- Experimental results show that the proposed method can effectively detect insulator flashover and breakage defects and reduce missing detection.

The rest of this paper is organized as follows. [Section 2](#) reports the related work. [Section 3](#) describes the details of the proposed methodology for defect detection in power lines. Experiment results on the defect detection are presented in [Section 4](#), and a short conclusion is finally drawn in [Section 5](#).

2 Related Work

Vision foundation models. Currently, the insulator defect detection models are mainly categorized into traditional image processing-based methods and deep learning-based methods. Methods based on traditional image processing design algorithms that involve the extraction of features [8], including color [9,10], morphology [11,12], gradient [13], edge [14], texture [15], and spatial characteristics [16]. The extracted features are poorly generalized to different tasks or objects and thus are gradually replaced by deep learning methods. Further, the methods are also susceptible to the interference of complex backgrounds, which is not conducive to small object detection.

The rise of deep learning has positioned it as a prominent technique in the intelligent inspection of power lines. The deep learning-based methods are categorized into two-stage methods and single-stage methods. Algorithms like R-CNN [17], Fast R-CNN [18], and Faster R-CNN [19] exemplify the two-stage methods. Reference [20] proposed an improved Faster R-CNN model based on deep learning to improve the precision of fault detection. The method first replaced the feature extraction network and used a feature pyramid for feature fusion, and finally used RoIAlign instead of the RoIPooling network to reduce the impact of quantization. Thus, the reduction of missed detection rate and false detection rate were realized. While these two-stage methods offer high detection precision, they face challenges in meeting real-time requirements in practical application scenarios. This is due to the candidate frame generation phase, which introduces significant computational redundancy, leading to a reduction in detection speed.

Single-stage methods, represented by algorithms such as the YOLO [21], YOLOv4 [22], YOLOv5 [23], YOLOv7 [24], and SSD [25], can directly predict the location and object class by using the location information as a potential object. Especially, the YOLO algorithm achieves a faster detection speed compared to two-stage methods, while taking into account a higher detection accuracy. Therefore, the YOLO series is widely used in industrial applications. Hao et al. [26] carried out a new architectural design from the backbone network and neck of YOLOv4, respectively. They designed CSP-ResNeSt to extract stronger features to weaken the influence of complex backgrounds in aerial images. Subsequently, Bi-SimAM-FPN, featuring split-attention blocks, was introduced to address the challenge of accurately identifying small-scale insulator defects. Reference [27] introduced Mina-Net for detecting self-blast in insulators, leveraging the YOLOv4 framework. The approach primarily incorporated shallow feature mapping within the feature pyramid and subsequently enhanced Squeeze-and-Excitation Networks (SENet) to recalibrate features across different levels in the channel direction. In another work, Ding et al. [28] integrated the Assumption-free K-MC2 (AFK-MC2) algorithm into YOLOv5, adapting the K-means method to enhance both accuracy and speed in detecting defects in insulator strings.

The architectural framework of the YOLO series primarily comprises three components, which are the backbone, neck, and head. In terms of extracting multi-scale features, top-down paths, and transversally connected paths can represent multi-scale objects correctly. However, this leads to the fact

that the different scale feature layers are only responsible for detecting objects at the corresponding scales, and the different scales are not sufficiently fused. In addition, tiny object detection still suffers from insufficient precision.

In this paper, we choose YOLOv5 as a baseline for four reasons: (1) YOLOv5 is the first model in the YOLO family to apply the gradient shunting idea to design a more efficient network architecture by using Cross Stage Partial Network (CSPNet) [29]. The subsequent YOLOv7 and YOLOv8 are both based on this idea to obtain richer gradient information by branching more gradient flow in parallel and thus obtain higher precision. Therefore, the value of using YOLOv5 as a baseline for our research results can be demonstrated more clearly in essence. (2) YOLOv5 is a popular and mature object detection model for industrial applications. It has been highly optimized compared to other versions and therefore has a very stable performance. (3) YOLOv5 has a strong advantage in the rapid deployment of the model. It not only dramatically improves detection speed while maintaining precision, but is also more friendly and flexible to deploy. (4) YOLOv5 has excellent FPS, and compared with YOLOv8, it is more suitable for deployment and real-time application on devices that do not support GPU. Besides, YOLOv7 and YOLOv8 all have higher parameters and FLOPs than YOLOv5 in the same level of model.

Attention mechanisms. In recent years, attention mechanisms have been widely used in the field of power line defect detection. Chen et al. [30] added SENet, a channel attention mechanism, to the YOLOv5 backbone network to improve the feature extraction ability of the model. In addition, some researchers combine SENet with other attention mechanisms to solve the problem of low accuracy of power line defect detection. For example, Efficient Channel Attention (ECA) and SENet formed a double attention fusion module [31]. Alternatively, SENet and Concentration-Based Attention Module (CBAM) were introduced respectively to merge object features at different scales and prominent feature information [32]. Transformer [33] based on the self-attention mechanism has achieved great success in the field of natural language processing and some scholars have also applied it in the field of power line defect detection [34,35]. However, Self-attention is computationally heavy because it processes all the pixels in an image. In contrast to the above methods, inspired by the Deformable Attention Mechanism (DAM) proposed in Deformable DETR [36], EDAM proposed in this paper makes use of DAM that can significantly reduce computation and have better performance than the self-attention mechanism. EDAM has more powerful attention ability than DAM (especially on small objects) and more stable training gradients due to the introduction of fusion gating activation.

3 Methodology

In this paper, we present a simple and effective surface defect detection method for power line insulators. This method addresses challenges such as low precision and missed detection caused by small defect objects, variable scales, and fuzzy edge morphology.

Fig. 1 illustrates the overall framework of the defect detection method proposed in this paper. The CBL is composed of convolution, batch normalization, and the Leaky ReLU activation function. There are two structures contained in the model, the CSP1_X structure and the CSP2_X structure, where X denotes multiple residual units. The CSP2 refers to $X = 1$, indicating that it contains only one residual unit. To maximize the retention of detailed features conducive to small object detection, a shallow high-resolution feature map of size $160 \times 160 \times 64$ in the backbone network is introduced firstly to be fused with the Feature Pyramid Network (FPN). Secondly, we introduce the ASFF module and improve it. The cross-scale fusion of FPN layers at each scale is performed by the adaptive spatial feature fusion layer S-ASFF, so that the fusion is dominated by the most important feature layer

at each spatial location, improving the multi-scale characterization capability. Finally, an enhanced deformable attention mechanism is introduced with a fusion gating activation function, which further inspires the model to learn a small number of critical sampling points in a set of sampling points, reducing the defect missed detection rate of weak edge features.

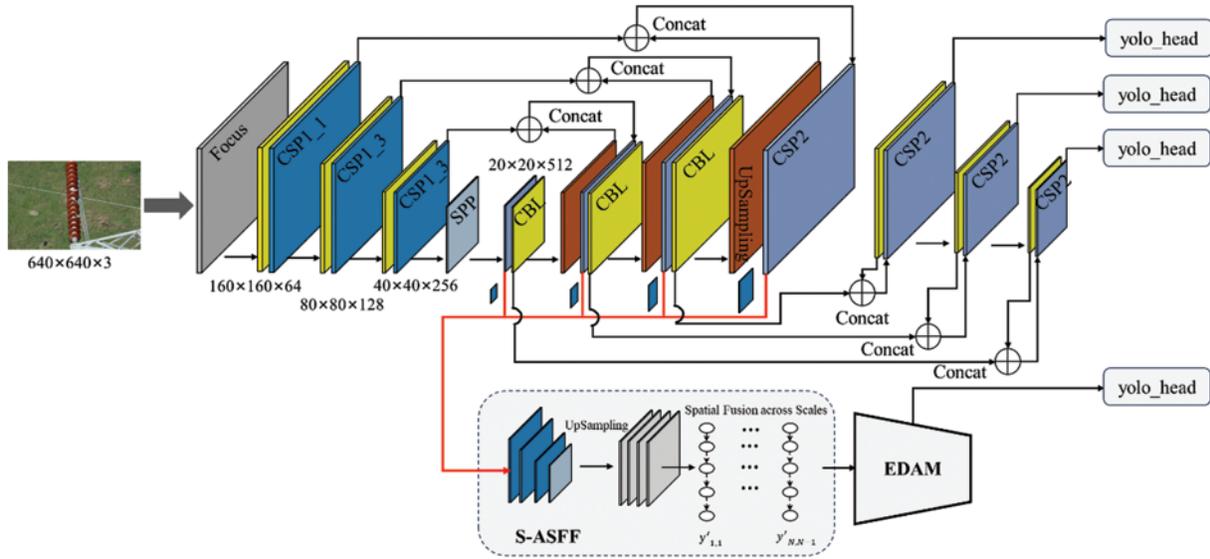


Figure 1: The overall architecture of the proposed method

3.1 Strategy Adjustment of YOLOv5 Network Structure

The following is the setting of FPN used to construct the feature pyramid. The original feature representations are denoted as $P = \{P_3, P_4, P_5\}$, which correspond to the multilevel feature maps $C = \{C_3, C_4, C_5\}$ with predetermined strides $\{8, 16, 32\}$ in feature hierarchy of the input image. To address the challenges that lead to defect missed detection due to insulators with small defect objects, low proportion of pixels occupied, and serious loss of detailed features during convolution down samplings, we introduce the high-resolution feature map C_2 after two down samplings in the backbone network and fuse it with P_2 of FPN. This way can enhance the detailed features of small objects in the neck network and alleviate the situation of serious loss of the detailed features in the deeper layers. The specific steps are: (1) P_3 is upsampled and then feature fused with C_2 to obtain P_2 which increases the detail features in the neck network that are favorable for small object detection; (2) the feature map P_2 is input into the YOLOv5 head after convolution, batch normalization, and Leaky ReLU activation, which ultimately improves the detection of tiny objects.

3.2 Simplified Adaptive Spatial Feature Fusion

To tackle the issue of insufficient feature fusion across different scales, which hampers the network of ability to adapt to significant variations in object scales, S-ASFF is added to the neck network, inspired by the Adaptive Spatial Feature Fusion (ASFF) [37] module. In addition, to reduce the number of parameters, we remove the feature fusion layers of ASFF-1, ASFF-2, and ASFF-3. By utilizing the feature fusion mechanism of S-ASFF, the model can adaptively learn the weights of the different scale feature layers of the FPN at each same position, so that the most important feature layers dominate the fusion. The comparison between the ASFF module and the S-ASFF module is

shown in Fig. 2. The original ASFF contains three fusion layers while S-ASFF has only one fusion layer which integrates the feature layer P_2 of FPN with more detailed features. The S-ASFF can fuse the bottom feature maps with higher resolution in the FPN so that it can improve the detection ability of smaller objects.

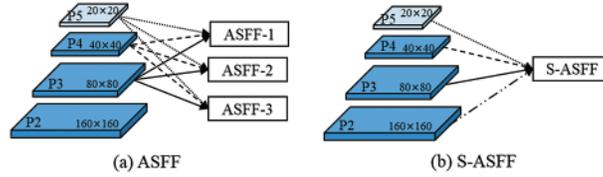


Figure 2: The network structure comparison between ASFF and S-ASFF

As shown in Fig. 3, S-ASFF is specifically divided into three steps: (1) up-sampling $P = \{P_3, P_4, P_5\}$ to expand the scale to ensure that the scales of each feature layer are consistent in the process of fusion; (2) spatial filtering for each position after up-sampling to learn the relative importance of different scale layers and to enhance the perceptibility of the scales; and (3) spatial fusion across scales to enhance the adaptability to changes in scales. As shown in Fig. 3, where P_5 is the highest layer and P_2 is the lowest layer. After up-sampling, the feature maps of each layer are denoted as x^k , $k \in [2, 5]$.

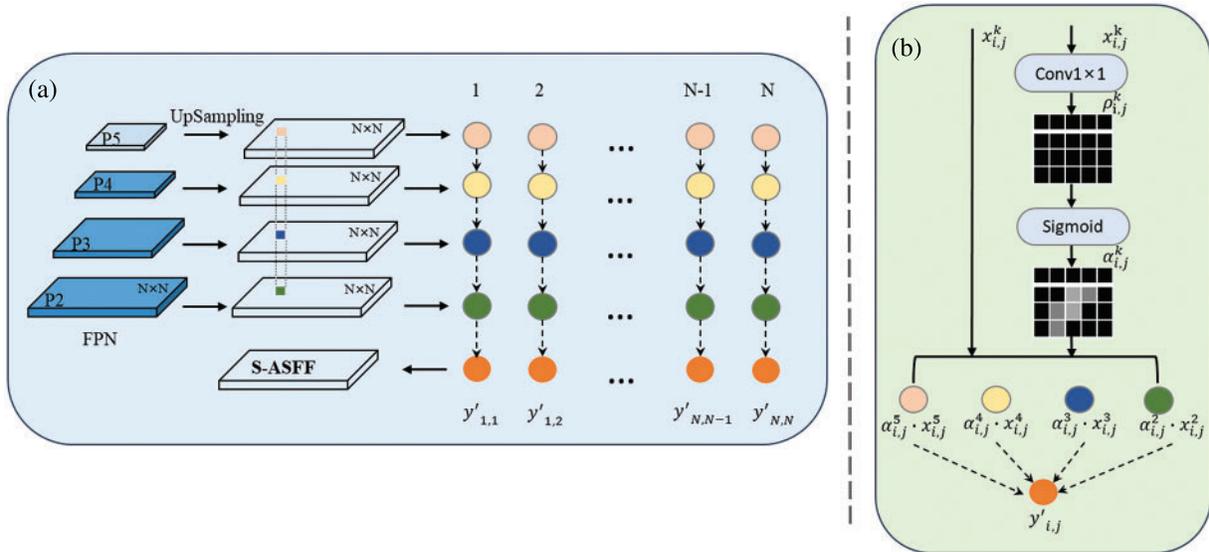


Figure 3: The structure of simplified adaptive spatial feature fusion. (a) Represents the overall architecture of S-ASFF with FPN. (b) Represents the details of spatial filtering and cross-scale fusion for S-ASFF

Both scale dimension and spatial dimension are taken into account in S-ASFF which processes spatial weights to the scaled feature maps at each level. A softmax activation function with a control factor ρ is used to calculate the spatial mask indicating the relative importance of the corresponding

perceptual domain of attention, extracting effective discriminant features. The formula is represented as follows:

$$\varnothing(x) = \max\left(0, \frac{\tanh(x - \varphi) + \tanh(\varphi)}{1 + \tanh(\varphi)}\right) \in [0,1], \quad (3)$$

where x is the feature map after convolution, φ is a predefined hyperparameter and \tanh is the hyperbolic tangent function.

Based on the features in the across-scale fusion space, the output after attention for K sampling points on a pixel p_0 is represented as follows:

$$\dot{F}(p_0) = \sum_{k=1}^K A_k \cdot w_k \cdot x(p_0 + p_k + \Delta p_k), \quad (4)$$

where \dot{F} is the feature space for EDAM. For a pixel p_0 on the feature map y after general convolution, K is the total number of sampling points on a pixel p_0 . We set $K = |\mathcal{R}|$. The range of the convolution kernel is defined by \mathcal{R} . In this paper, $\mathcal{R} = \{((-1, -1), (-1, 0), \dots, (0, 1), (1, 1))\}$, which is represented that 3×3 convolution kernel with dilation 1. A_k is a mask value for the k -th sampling point by gating activation function \varnothing . $p_k + \Delta p_k$ is the offset position of the k -th sampling point. It is worth noting that when the offset coordinates are floating point numbers, we use the nearest neighbor interpolation method to round the offset coordinates to obtain the revised coordinates and determine the offset position $p_k + \Delta p_k$.

4 Experiments

4.1 Dataset Preparation

This paper utilizes a dataset of 1600 raw images. We first collated insulator images containing flashover and breakage from the Electric Power Research Institute (EPRI) and public datasets UPID [38,39]. And then we performed a pre-processing operation in images: Cleaning up the damaged images and resizing them. In addition, image flipping, saturation adjustment, contrast adjustment, and noise addition were used to expand the data. Second, a dataset in YOLO format was produced for training and evaluation. Specifically, the images were labeled using the Labeling tool to obtain the object categories and coordinate information of ground truth. There are three object classes: Pollution_flashover, broken, and insulator. According to statistics, the sample size of these three classes is 1994, 861, and 1466 in order. The dataset was partitioned into a training set and a test set, with an 8:2 ratio.

The label correlogram for objects of each size in the dataset is shown in Fig. 5. It can help identify patterns or correlations in the distribution of object annotations across different classes and scales. It reveals if certain classes are more likely to appear at specific scales [40]. x and y are the coordinates of the bounding box of the object, and width and height are the size of the bounding box of the object, respectively. From the three coordinate graphs with red labeled boxes, it can be shown that most of the bounding box widths and heights are less than 1/4 or 1/8 of the original size of the entire image. However, because the size of the feeding neural network is fixed at 640×640 , most of the object pixels in the actual trained images take up a smaller proportion. Overall, the object bounding boxes are distributed in all parts of the x-axis, indicating that the object scale varies greatly. In addition, this dataset covers most of the defect scenes in practical applications which makes the trained model generalizable. To further reduce the network overfitting problem, the model was enhanced with a mosaic data enhancement method for the samples in the dataset before training. Furthermore, data

enhancement strategies such as random scaling and random cropping were used to improve the model classification performance.

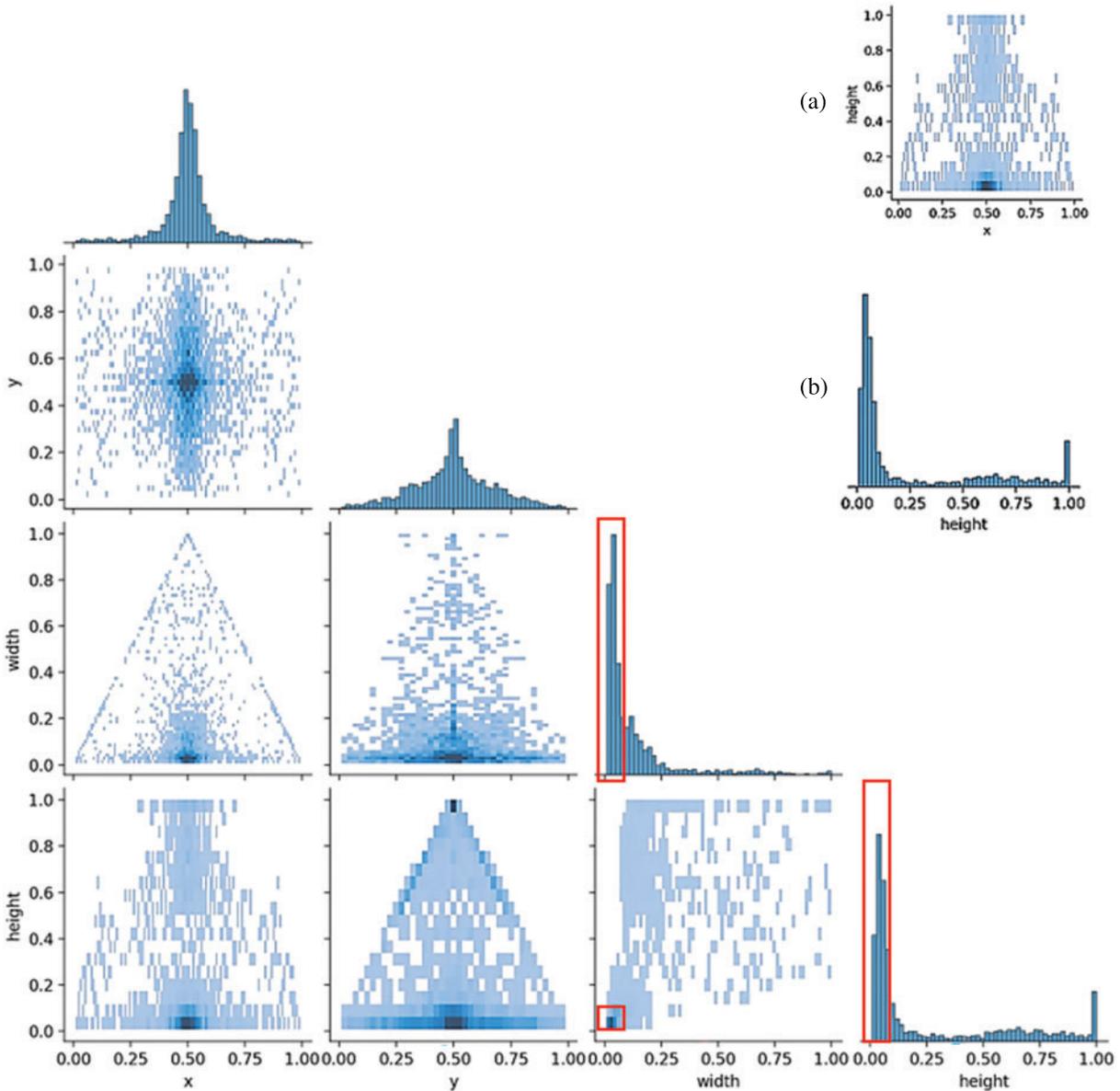


Figure 5: The label correlogram for objects of each size in the dataset. (a) Represents a class of graphs depicting the relationship between x, y, width, and height. Similarly, (b) represents a class of graphs depicting the distribution

4.2 Experimental Metrics and Implementation Details

For a more comprehensive evaluation of the model, the paper employed three metrics: Precision (P), Recall (R), and F1-Score, to account for the comprehensive prediction of breakage and flashover. Additionally, mAP was utilized as an overarching performance metric to characterize and assess the

model's quality. All the experiments in this paper were conducted in the hardware environment shown in [Table 1](#).

Table 1: The experimental running environment

Type	Configuration
CPU	Intel(R) Core(TM) i5-10500 CPU @ 3.10 GHz
GPU	Nvidia GeForce RTX3090 (24 G)
Accelerated environment	CUDA 11.5
Operating system	Windows10
Deep learning framework	Pytorch1.10.1
Programming language	Python3.7.12

4.3 Model Training

The network was fed with images of size 640×640 , and several key parameters were configured. The batch size was set to 8, momentum to 0.937, initial learning rate to 0.01, and weight decay to 0.0005. The network was trained from scratch for 300 epochs. For the constructed dataset, the number of object classes is inevitably unbalanced. To alleviate the impact of this problem, we adopted YOLOv5 with a class imbalance strategy as the baseline. This strategy is proposed by YOLOv5. The setting of class weights and image weights was introduced. The class weights require calculating the number of labels for each class in the dataset and then taking the reciprocal of the number of class labels. In other words, the greater the number of labels of a certain class, the smaller the weight in the image containing that class. Calculating the sum of all the class weights contained in an image is the image weight. The greater the image weight, the greater the probability of the image being sampled. In particular, the image is selected according to the image weight by random selection method and the number of images is the same as that of the training set. This means that the larger image weight will be more likely to be selected for training. This method can increase the training proportion of the small number of object classes during training so that the model can learn the features of each object class more balanced and prevent over-fitting. The training results are depicted in [Fig. 6](#). As the number of training epochs gradually increased to 300, the loss curves stabilized, indicating effective training.

4.4 Comparison with the SOTA Models

In this section, we assessed the detection capabilities of our method. For a fair comparison, we re-implemented the relevant models and employed the same evaluation dataset to calculate performance metrics, including precision, recall, and F1-Score, for the different models. All results are presented in [Table 2](#). The detection model based on Faster RCNN and SSD has low detection precision and recall for flashover and breakage as small objects due to the limitations of the basic network framework, so they are difficult to apply in practice. Compared with YOLOv5s, both YOLOv7 and YOLOv8s have better performance. However, when applied to YOLOv5s, YOLOv7, and YOLOv8s, respectively, our proposed method can bring them greater precision gains and recall rate gains. In particular, when the YOLOv5s baseline was combined with our proposed method, the precision on flashover and breakage defects was improved by 5.7% and 4.7%, respectively. This exceeds the YOLOv8s baseline, even the model YOLOv8s + Ours, and significantly narrows the performance gap with YOLOv7. As for the higher gains of our method on YOLOv5, the possible reason is that the YOLOv5 model is not robust

enough. There are more robust network structures and performance due to the optimization and improvement of YOLOv7 and YOLOv8 in the aspects of model feature extraction and strategy for matching positive and negative samples.

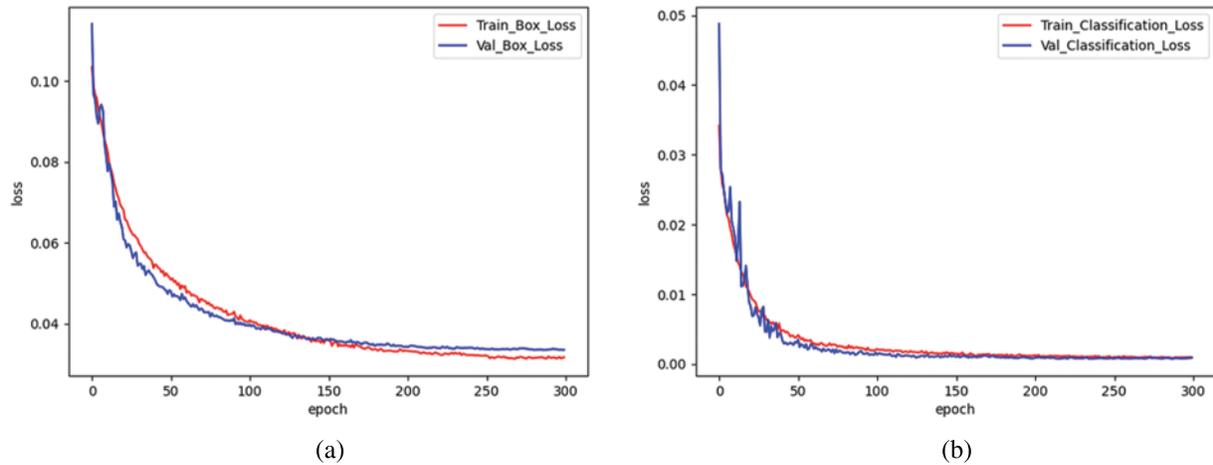


Figure 6: The graphs of training loss and validation loss. (a) Illustrates the bounding box regression loss. (b) Represents the classification loss

Table 2: The comparison with the state-of-the-art models

Model	Breakage (%)			Flashover (%)			mAP
	P	R	F1-Score	P	R	F1-Score	
Faster RCNN	60.6	81.5	69.0	43.6	60.8	51.0	62.8
SSD	91.2	75.4	83.0	74.9	67.6	71.1	75.3
FINet [41]	90.3	87.5	88.9	87.3	78.5	82.7	80.5
AFNet	88.0	87.5	87.7	88.1	81.5	84.6	81.3
YOLOv5s	87.3	89.6	88.4	83.4	83.0	83.2	83.5
YOLOv5s + Ours	93.1 + 5.7	90.6 + 1.0	91.8	88.1 + 4.7	86.0 + 3.0	87.0	85.6 + 2.1
YOLOv7	94.7	91.3	93.0	89.8	86.3	88.0	87.4
YOLOv7 + Ours	95.4 + 0.7	92.7 + 1.4	94.0	92.0 + 1.2	87.2 + 0.9	89.5	87.9 + 0.5
YOLOv8s	90.3	89.6	89.9	87.5	75.7	81.2	84.9
YOLOv8s + Ours	92.2 + 1.9	89.9 + 0.3	91.0	88.0 + 0.5	79.4 + 3.7	83.5	86.1 + 1.2

We also provide the deployment performance of each model in practical applications, as shown in Table 3. In the task of power line inspection, the deployment of deep learning models is a comprehensive process involving several key factors. How efficiently the model is deployed on specific hardware for optimal performance is critical. YOLOv7 and YOLOv8s are higher than YOLOv5s

in FLOPs, Parameters, and Model Size. For deploying models on hardware devices that do not support GPU, YOLOv5s is more suitable. However, the YOLOv5s is significantly inferior to the YOLOv5s + Ours in terms of precision. In actual model deployment, compared with YOLOv7 + Ours and YOLOv8s + Ours, YOLOv5s + Ours has the advantages of small model size, high precision, low computational efficiency and flexible deployment. If optimal precision needs to be considered, YOLOv7 as the baseline can be chosen to detect flashover and breakage defects of power line insulators.

Table 3: The comparison of deployment performance in practical application

Model	FLOPs	Parameters (MB)	Model size (MB)
Faster RCNN	474.0	28.8	108.2
SSD	137.0	24.0	91.6
YOLOv5s	17.0	7.3	14.1
FINet	16.9	7.3	14.2
AFNet	17.7	8.1	15.8
YOLOv7	53.2	37.6	71.3
YOLOv8s	28.7	11.2	21.5
YOLOv5 + ours	23.1	8.4	16.6

The results show that our proposed method is effective and lightweight in solving small objects, scale variations, and weak edge morphological features. Firstly, for strategy adjustment of the YOLOv5 network structure, although this strategy only uses simply high-resolution feature map, it brings two major improvement advantages to the model: (1) reduces the loss of details of small objects; (2) brings smaller object-scale information to the second component (S-ASFF). This balances the preference for learning large and small objects, allowing the model to learn tiny object features that are more difficult to learn. Secondly, for S-ASFF, the improved performance gains benefit from the fact that the most important scale layer dominates each position in the feature pyramid with minimal computational cost. For better fusion, we weigh the scale to which the feature pyramid should be scaled with the least computational cost while preserving as much of the small object information as possible from the first proposed component (strategy). Other fusion ways all cause secondary loss of small object information or increase unnecessary computing costs. Finally, due to EDAM introducing the gating activation function \varnothing with tanh instead of the sigmoid in DAM, EDAM can suppress all attention values with negative input to 0, which is similar to the negative half region of the ReLU. This will increase the sparsity of attention. Compared to DAM which indiscriminately learns from a small group of sampling points near a certain pixel point, EDAM can focus on more meaningful location learning discriminant features, resulting in improved performance. In addition, EDAM alleviates gradient dispersion because the sensitive range of gradient change is expanded from [0,0.25] to [0,1], and the gradient is more stable during training. Furthermore, our method is a more generic variant method compared to the existing solutions on the YOLO series, which can improve the performance of models in the YOLO family.

All in all, in this paper, a general enhancement method is designed, which reasonably uses high-resolution feature maps and can simply and effectively improve the adaptability of variable multi-scale features and the perception of weak edge features with less increase of parameters and computation. It significantly enhances the detection performance of the YOLO series on difficult small objects. The missing detection of defects in power lines is reduced.

To verify the detection performance of our model on large-scale objects and make the results more convincing, we trained and tested various mainstream algorithms on the public dataset UPID and compared them with the IFD dataset constructed in this paper. As shown in Table 4. For large-scale objects, the AP of our model on both the IFD dataset and the UPID public dataset is best.

Table 4: The performance comparison on different datasets

Model	IFD	UPID
	Insulator AP (%)	Insulator AP (%)
SSD	97.46	91.21
YOLOv5s	97.50	93.74
FINet	98.39	94.59
AFNet	98.21	94.95
Ours	98.48	95.45

Object detection was performed on the test set and the results were visualized. Fig. 7 shows some of the results. The experiment proves that the detection model has strong generalization capability and wide application potential.

4.5 Ablation Study

In this section, we conducted essential ablation experiments to quantitatively analyze the effectiveness of the various equipment proposed. We gradually introduced the relevant modules into the network during training. YOLOv5s served as the baseline method for all ablation studies. For convenience, the strategy adjustment of YOLOv5s network structure is denoted as SAStgy. The results are shown in Table 5.

We note that directly adding an SAStgy makes slight degradation for the performance of breakage but improves the precision of flashover. The result suggests an unstable effect. We consider possible reasons for this: The plain YOLOv5s has no advantage for the detection of small objects due to variable scale and weak edge features besides complex backgrounds. The method has difficulty in precise judgment for the defect location and damage degree. For specific defect detection tasks, further optimized strategies are necessary to improve detection performance.

The results show that S-ASFF and EDAM, based on the addition of a SAStgy, boost the performance of the baseline: The detection precision of flashover is improved by 3.01% and that of breakage by 4.58%. However, EDAM improves the model performance most significantly. The detection precision of flashover is boosted by 4.71% and recall by 2.96% over YOLOv5s. The detection precision of breakage is increased by 5.74% and recall by 1.05%. This is because S-ASFF provides key details to EDAM and reduces interference from complex backgrounds. It enables EDAM to more accurately expand the perceptual domain of attention and extract discriminative features.

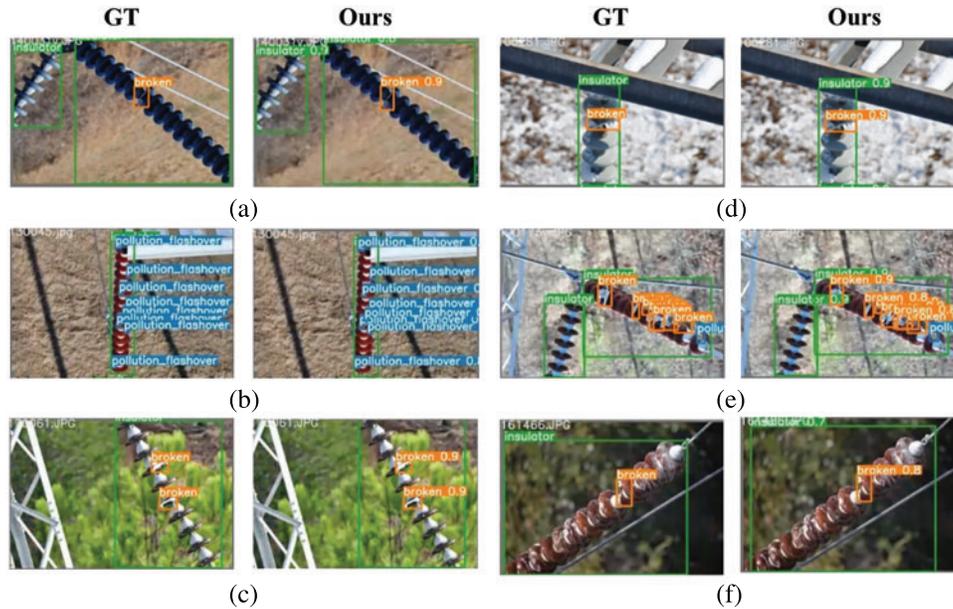


Figure 7: The visualization of defect detection results of power line equipment. (a) Illustrates the objects can be effectively identified when the attitude is deformed. (b) Illustrates dense and weakly discharged flashovers can also be effectively detected. (c) Represents broken cross-sections can also be effectively detected when the light is intense. (d) Represents detection results for ice-covered scenarios. (e) Illustrates no missed detection for the case of dense breakage defects and overlapping edges on an insulator. (f) Severe accumulation of dirt on insulators is not mistakenly detected as a flashover

Table 5: The effect of each component

Model	Breakage (%)			Flashover (%)		
	P	R	F1-Score	P	R	F1-Score
YOLOv5s	87.31	89.58	88.43	83.37	83.04	83.20
YOLOv5s + SAStgry	86.92	89.10	88.00	86.77	85.40	86.08
YOLOv5s + SAStgry + S-ASFF	91.89	88.54	90.19	86.45	85.60	86.03
YOLOv5s + SAStgry + S-ASFF + DAM	92.38	88.85	90.58	86.95	85.02	85.97
YOLOv5s + SAStgry + S-ASFF + EDAM	93.05	90.63	91.82	88.08	86.00	87.03

To illustrate the benefits of our model more visually. This is shown in Fig. 8. The first row shows an example of breakage and the second row shows an example of flashover. The layer23 will be further used to predict small objects. It is used together with layer 26 and layer 29 as input to yolo heads for classification and regression tasks. Before S-ASFF (as shown in layer 21), features are retained in the object regions because of the stronger semantic information. On the contrary, after the object scale perception of S-ASFF (as shown in layer 22), features are extracted around each object. Compared with layer 21, large-scale insulators and small-scale defect (flashovers and breakages)

objects significantly improve the feature distinction in scale and are more sensitive to location. After EDAM (as shown in layer 23), distinguishing features are highlighted from key information retained by the previous module, and object sizes and edge morphological features are precisely learned. At the same time, one may notice that the defect objects can be learned in both layer 26 and layer 29 because YOLOv5 allows ground truth boxes to perform anchor matching in all prediction layers simultaneously to increase the number of positive samples. However, layer 23 contains more accurate small object size and richer edge morphology. Therefore, for SAStgy, adding a tiny object prediction layer to directly predict the feature map of layer23 is conducive to improving the precision of smaller objects, especially for detecting difficult small objects.

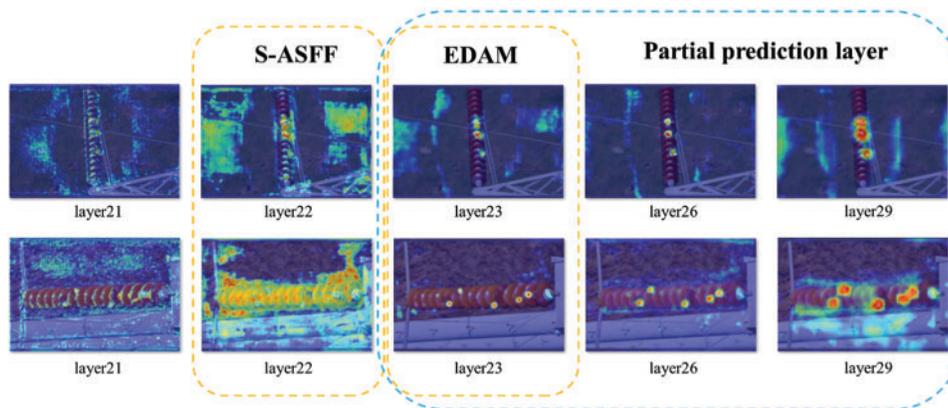


Figure 8: The results of heat map visualization

4.6 Runtime Analysis

We also recorded the testing time of our method based on YOLOv5. Specifically, with a batch size of 16, our method achieved an inference time of 6.1 ms and a non-maximum suppression (NMS) time of 0.9 ms. Consequently, the total processing time for an image with a size of 640 pixels was 7.1 ms.

5 Conclusion

In this paper, we develop research on defect detection in power lines based on YOLOv5, a framework of object detection algorithms in the field of computer vision. We analyze the challenges in this field. Through image dataset construction, image processing, model improvement, and experimental validation of flashover and breakage, we propose a simple and effective surface defect detection method of power line insulators for difficult small objects. After experimental validation, we conclude that our model has the advantages of high precision, low omission rate, stability, and fast convergence when compared with state-of-the-art detection models. It can detect objects in real time. our model can be extended to other defect detection, such as bird's nests on power lines or towers, and hanging foreign objects. The fact that deep learning-based defect detection for flashover and breakage in power lines is under-reported in the literature, so this paper has clear engineering research value. If more sufficient and diverse defect image data can be obtained, the next step will focus on multiple defect detection scenes of various power components, and solve the difficult small object detection problem in this scene with multi-scale objects.

Acknowledgement: The authors would like to express their gratitude to State Grid Jiangsu Electric Power Co., Ltd. for support of this study.

Funding Statement: This research was funded by State Grid Jiangsu Electric Power Co., Ltd. of the Science and Technology Project (Grant No. J2022004).

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Xiao Lu, Chengling Jiang; data collection: Xiao Lu, Zhoujun Ma; analysis and interpretation of results: Xiao Lu, Chengling Jiang, Haitao Li; draft manuscript preparation: Xiao Lu, Chengling Jiang, Yuexin Liu. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The datasets can be openly available in UPID at <https://github.com/heitorcfelix/public-insulator-datasets>. The code/supplementary data supporting the findings of this study are available from the corresponding authors upon a request.

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

References

- [1] Y. J. Zhai, D. Wang, M. L. Zhang, J. R. Wang, and F. Guo, "Fault detection of insulator based on saliency and adaptive morphology," *Multimed. Tools Appl.*, vol. 76, no. 9, pp. 12051–12064, 2017. doi: [10.1007/s11042-016-3981-2](https://doi.org/10.1007/s11042-016-3981-2).
- [2] X. Mei, T. C. Lu, X. Y. Wu, and B. Zhang, "Insulator surface dirt image detection technology based on improved watershed algorithm," in *Proc. 2012 Asia-Pacific Power and Energy Engineering Conf.*, Shanghai, China, 2012, pp. 1–5.
- [3] M. D. F. Ahmed, J. C. Mohanta, and A. Sanyal, "Inspection and identification of transmission line insulator breakdown based on deep learning using aerial images," *Electr. Power Syst. Res.*, vol. 211, no. 5, pp. 108199, 2022. doi: [10.1016/j.epsr.2022.108199](https://doi.org/10.1016/j.epsr.2022.108199).
- [4] L. Zhang, B. Q. Li, Y. Cui, Y. S. Lai, and J. Gao, "Research on improved YOLOv8 algorithm for insulator defect detection," 2023. Accessed: 15 Sep. 2023. [Online]. Available: <https://doi.org/10.21203/rs.3.rs-3337929/v1>
- [5] X. T. Zhang *et al.*, "InsuDet: A fault detection method for insulators of overhead transmission lines using convolutional neural networks," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021. doi: [10.1109/TIM.2021.3127641](https://doi.org/10.1109/TIM.2021.3127641).
- [6] H. Jiang, "Multi-defect detection of transmission line insulators based on YOLOV5 UAV aerial photography," M.S. dissertation, Guangdong Univ. of Tech., China, 2021.
- [7] H. Xu, "Research on small target detection and defect identification of transmission lines based on machine vision," M.S. dissertation, Zhejiang Univ., China, 2022.
- [8] H. Wang, L. Cheng, R. Liao, S. Zhang, and L. Yang, "Nonlinear mechanical model of composite insulator interface and nondestructive testing method for weak bonding defects," *Chin. J. Electr. Eng.*, vol. 39, no. 3, pp. 895–905, 2019. doi: [10.1049/hve.2019.0044](https://doi.org/10.1049/hve.2019.0044).
- [9] Y. P. Ma, Q. W. Li, L. L. Chu, Y. Q. Zhou, and C. Xu, "Real-time detection and spatial localization of insulators for UAV inspection based on binocular stereo vision," *Remote Sens.*, vol. 13, no. 2, pp. 230, 2021. doi: [10.3390/rs13020230](https://doi.org/10.3390/rs13020230).
- [10] S. Y. Ma, J. B. An, and F. U. Chen, "Segmentation of insulator images based on HIS color space," *Journal of Dalian Nationalities University*, vol. 12, no. 5, pp. 481–484, 2010.
- [11] Y. T. 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," *Electronic Power*, vol. 50, no. 5, pp. 52–58, 2017.

- [12] Q. Wu and J. An, "An active contour model based on texture distribution for extracting inhomogeneous insulators from aerial images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3613–3626, 2014. doi: [10.1109/TGRS.2013.2274101](https://doi.org/10.1109/TGRS.2013.2274101).
- [13] H. B. Zai, L. He, and Y. F. Liu, "Target tracking method of transmission line insulator based on multi feature fusion and adaptive scale filter," in *Proc. ACPEE*, Chengdu, China, 2020, pp. 4–7.
- [14] J. G. Yin, Y. F. Lu, Z. X. Gong, Y. C. Jiang, and J. G. Yao, "Edge detection of high-voltage porcelain insulators in infrared image using dual parity morphological gradients," *IEEE Access*, vol. 7, pp. 32728–32734, 2019. doi: [10.1109/ACCESS.2019.2900658](https://doi.org/10.1109/ACCESS.2019.2900658).
- [15] B. F. Li, D. L. Wu, Y. Cong, Y. Xia, and Y. D. Tang, "A method of insulator detection from video sequence," in *Proc. 2012 Fourth Int. Symp. Eng. Educ.*, Shanghai, China, 2012, pp. 386–389.
- [16] H. Y. Cheng, Y. J. Zhai, R. Chen, D. Wang, Z. Dong and Y. T. Wang, "Self-shattering defect detection of glass insulators based on spatial features," *Energies*, vol. 12, no. 3, pp. 543, 2019. doi: [10.3390/en12030543](https://doi.org/10.3390/en12030543).
- [17] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. CVPR*, Columbus, OH, USA, 2014, pp. 580–587.
- [18] R. Girshick, "Fast R-CNN," in *Proc. CVPR*, Boston, MA, USA, 2015, pp. 1440–1448.
- [19] S. Q. Ren, K. M. He, R. Girshick, and J. Sun, "Faster R-CNN towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017. doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031).
- [20] J. P. Tang, J. Wang, H. L. Wang, J. Y. Wei, Y. J. Wei and M. S. Qin, "Insulator defect detection based on improved Faster R-CNN," in *Proc. AEEES*, Chengdu, China, 2022, pp. 541–546.
- [21] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. CVPR*, Las Vegas, NV, USA, 2016, pp. 779–788.
- [22] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," arXiv:2004.10934, 2020.
- [23] G. Jocher *et al.*, "ultralytics/yolov5:v6.0-YOLOv5n 'Nano' models, Roboflow integration, TensorFlow export, OpenCV DNN support," *Zenodo*, 2021.
- [24] C. Y. Wang, A. Bochkovskiy, and H. Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. CVPR*, Vancouver, Canada, 2023, pp. 7464–7475.
- [25] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Proc. ECCV*, Cham, Berlin, 2016, pp. 21–37.
- [26] K. Hao, G. K. Chen, L. Zhao, Z. S. Li, Y. L. Liu and C. Q. 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. doi: [10.1109/TIM.2022.3200861](https://doi.org/10.1109/TIM.2022.3200861).
- [27] H. He *et al.*, "An insulator self-blast detection method based on YOLOv4 with aerial images," *Energy Rep.*, vol. 8, no. 2018, pp. 448–454, 2022. doi: [10.1016/j.egy.2021.11.115](https://doi.org/10.1016/j.egy.2021.11.115).
- [28] J. Ding, H. N. Cao, X. L. Ding, and C. H. An, "High accuracy real-time insulator string defect detection method based on improved YOLOv5," *Front. Energy Res.*, vol. 10, pp. 928164, 2022. doi: [10.3389/fenrg.2022.928164](https://doi.org/10.3389/fenrg.2022.928164).
- [29] C. Y. Wang, H. Y. Mark Liao, Y. H. Wu, P. Y. Chen, and J. W. Hsieh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proc. CVPRW*, Seattle, WA, USA, 2020, pp. 1571–1580.
- [30] J. L. Chen, Z. J. Fu, X. Chen, and F. Wang, "A method for power lines insulator defect detection with attention feedback and double spatial pyramid," *Electr. Power Syst. Res.*, vol. 218, no. 99, pp. 109175, 2023. doi: [10.1016/j.epsr.2023.109175](https://doi.org/10.1016/j.epsr.2023.109175).
- [31] L. R. Li *et al.*, "Insulator defect detection based on multi-scale feature coding and dual attention fusion," *Advances in Laser and Optoelectronics*, vol. 59, no. 24, pp. 1–10, 2022.
- [32] J. Xie, Y. W. Du, Z. J. Liu, H. Liu, T. Y. Wang and M. Mou, "Visible light insulator defect detection algorithm based on lightweight improved YOLOv5s," 2022. Accessed: 19 Nov. 2022. [Online]. Available: <https://doi.org/10.13335/j.10003673.pst.2022.1438>
- [33] A. Vaswani *et al.*, "Attention is all you need," in *Proc. 31st Int. Conf. on Neural Information Processing Systems*, Long Beach, CA, USA, 2017, pp. 6000–6010.

- [34] Y. Z. Cheng, J. F. Zhang, and B. Y. Sun, "Defect detection model of Angle tower bolts based on Transformer and attention mechanism," *Computer System Application*, vol. 32, no. 4, pp. 248–254, 2023.
- [35] A. Xue, E. Y. Jiang, W. T. Zhang, S. F. Lin, and Y. Mi, "Foreign body detection in transmission line channel based on the fusion of window self-attention network and YOLOv5," *Journal of Shanghai Jiaotong University*, 2023. Accessed: 30 Nov. 2023. [Online]. Available: <https://doi.org/10.16183/j.cnki.jsjtu.2023.301>
- [36] X. Z. Zhu, W. J. Su, L. W. Lu, B. Li, X. G. Wang and J. F. Dai, "Deformable DETR: Deformable transformers for end-to-end object detection," arXiv:2010.04159, 2021.
- [37] S. Liu, D. Huang, and Y. Wang, "Learning spatial fusion for single-shot object detection," arXiv:1911.09516, 2019.
- [38] M. J. B. Reddy, B. K. Chandra, and D. K. Mohanta, "A DOST based approach for the condition monitoring of 11 kV distribution line insulators," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 18, no. 2, pp. 588–595, 2011.
- [39] V. S. Andrel, T. Chaves, and H. Felix, "Unifying Public datasets for insulator detection and fault classification in electrical power lines," 2020. Accessed: 12 Mar. 2020. [Online]. Available: <https://github.com/heitorcfelix/public-insulator-datasets>
- [40] G. Jocher, "Explaining the labels_correlogram.jpg," 2023. Accessed: 12 Oct. 2021. [Online]. Available: <https://github.com/ultralytics/yolov5/issues/5138>
- [41] Z. D. Zhang *et al.*, "FINet: An insulator dataset and detection benchmark based on synthetic fog and improved YOLOv5," *IEEE Trans. Instrum. Meas.*, vol. 71, no. 8, pp. 1–8, 2022. doi: [10.1109/TIM.2022.3194909](https://doi.org/10.1109/TIM.2022.3194909).