algorithms

MDPI

Article

FocusNet: A Lightweight Insulator Defect Detection Network via First-Order Taylor Importance Assessment and Knowledge Distillation

Yurong Jing [1], Zhiyong Tao [1,*] and Sen Lin [2]

[1] School of Electronic and Information Engineering, Liaoning Technical University, Huludao 125105, China; 472320477@stu.lntu.edu.cn
[2] School of Automation and Electrical Engineering, Shenyang Ligong University, Shenyang 110168, China; linsen@sylu.edu.cn
* Correspondence: taozhiyong@lntu.edu.cn

Abstract

In the detection of small targets such as insulator defects and flashovers, the existing YOLOv11 has problems such as insufficient feature extraction and difficulty in balancing model lightweight and detection accuracy. We propose a lightweight architecture called FocusNet based on YOLOv11n. To improve the feature expression ability of small targets, Aggregation Diffusion Neck is designed to achieve deep integration and optimization of features at different levels through multiple rounds of multi-scale feature fusion and scale adaptation, and Focus module is introduced to focus on and strengthen the key features of small targets. On this basis, to achieve efficient deployment, the Group-Level First-Order Taylor Expansion Importance Assessment Method is proposed to eliminate channels that have little impact on detection accuracy to streamline the model structure. Then, Channel Distribution Distillation compensates for the slight accuracy loss caused by pruning, and finally achieves the dual optimization of high accuracy and high efficiency. Furthermore, we analyze the interpretability of FocusNet via heatmaps generated by KPCA-CAM. Experiments show that FocusNet achieves 98.50% precision and 99.20% mAP@0.5 on a proprietary insulator defect detection database created for this project using only 3.80 GFLOPs. This research provides reliable technical support for insulator monitoring in power systems.

Keywords: insulator; deep learning; YOLOv11; FocusNet; defect detection

check for updates

Academic Editor: Muhammad Adnan Khan

Received: 22 August 2025
Revised: 11 October 2025
Accepted: 14 October 2025
Published: 16 October 2025

Citation: Jing, Y.; Tao, Z.; Lin, S. FocusNet: A Lightweight Insulator Defect Detection Network via First-Order Taylor Importance Assessment and Knowledge Distillation. *Algorithms* **2025**, *18*, 649. https://doi.org/10.3390/a18100649

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Insulators are a key component of the power transmission and distribution system. They are widely used in ultra-high-voltage, high-voltage, and medium-voltage transmission and distribution lines. They are suspended on towers to support and suspend conductors. Their core function is to ensure that the current does not leak to the ground or tower when it is transmitted along the conductor, and to avoid short-circuit faults between conductors of different phases [1,2]. In addition to insulation performance, insulators must have excellent mechanical properties, such as tensile strength and bending strength, to withstand the combined stress caused by the conductor's own weight, ice load and wind force; at the same time, they must withstand extreme temperature changes, ultraviolet radiation and other harsh environmental conditions [3,4]. As a key barrier to the safe

operation of transmission lines, power plants and substations, the reliability of insulators directly affects the power supply stability of the power system.

Insulator electrical damage is mainly caused by surface pollution, such as industrial dust, salt spray, bird droppings, and other pollutants, which form conductive channels after being damp, which trigger leakage current and cause local heating, thus forming a dry zone and finally inducing surface flashover [5]. When ambient humidity is high, the water film on the surface of the insulator directly connects to the electrode and can also cause flashover failure. In addition, mechanical damage is caused mainly by long-term vibration fatigue or impact by external forces, while environmental factors can cause material aging and performance degradation [2]. Since insulator damage is usually gradual and hidden, early proactive detection is the key link to prevent major safety accidents and ensure the economic operation of the power system [6,7].

Computer vision technology has become the core means of intelligent insulator detection by building a closed-loop system of automated inspection, high precision identification and real-time decision making [8,9]. Unmanned aerial vehicles equipped with computer vision systems can efficiently scan transmission lines, especially in areas blind to manual inspections such as high voltage levels and complex terrain, significantly improving operational safety [10]. Although the Transformer architecture has shown excellent performance in target detection, its high computational complexity makes it difficult to meet the real-time drone detection requirements, and there is a risk of overfitting [11].

Feng et al. [12] combined the bidirectional feature pyramid network (BiFPN) with the attention mechanism, achieved a recall of 89.9% and a precision of 96.5% on a self-built dataset, and successfully deployed it on a drone platform. Ahmed et al. [13] constructed a medium-scale insulator dataset and achieved edge detection through a self-developed quadcopter system, with a detection precision of 93.5% and a processing speed of 58.2 FPS. Gouda et al. [3] developed a wireless intelligent device based on real-time leakage current monitoring and achieved pollution flashover prevention through a cloud-based early warning system; after 50 tests, the accuracy was 91.66%. These studies provide important technical support for the engineering application of intelligent insulator detection.

However, automated insulator inspection still faces many challenges [14,15], mainly in the following three aspects. To avoid electromagnetic interference, drones need to maintain a relatively safe distance from transmission lines during inspections. Therefore, in drone-captured image datasets, the target scale of insulator defects is usually relatively small, making detection more difficult. Unmanned aerial vehicles collect images from different angles during inspections, so the angles of insulators are relatively diverse. It is necessary to explore new algorithms to adapt to the angle changes of insulators. In automated insulator inspection, it is necessary to expand research on lightweight methods and explore new technical paths.

To address these issues, this work makes three contributions.

(1) Based on YOLOv11, we propose a new architecture called FocusNet that integrates Aggregation Diffusion Neck and Focus modules, aiming to solve the problems that key features are easily submerged by noise and multi-scale feature matching is misaligned during detection.

(2) To adapt to the computing power of edge devices, we combine the evaluation of the importance of Taylor expansion at the group level with the knowledge distillation of the channel distribution and apply them to the structurally improved detection framework. This achieves a lightweight model while maintaining the expression capabilities of key features.

(3) Combined with heatmaps generated by KPCA-CAM, we perform an interpretable analysis of the detection framework to intuitively reveal the key areas focused on by the model.

2. Related Work

Convolutional Neural Network (CNN) has shown unique advantages in insulator defect detection due to its translation equivariance of convolution kernels. Regardless of the position of the defect in the image, the convolution kernel with a specific feature response can be activated, and the shared application mechanism of the same convolution kernel in the entire image domain has become the core foundation for its efficient detection [16]. In the CNN hierarchical structure, the shallow network extracts basic features such as the edge texture of the insulator, the middle network realizes component-level feature combination, and the deep network completes semantic object recognition, thereby realizing the positioning and classification of the complete defect area [17].

Faster R-CNN is a typical two-stage object detection framework, which generates candidate regions through the Region Proposal Network and then performs classification and regression operations. Zhou et al. [18] incorporated the attention mechanism into Mask RCNN to pay more attention to small objects, with an average accuracy of up to 98%. Although the improved algorithm based on this architecture has been widely used in insulator detection scenarios, the two-stage processing flow leads to a significant real-time bottleneck [19]. In contrast, the YOLO series, as a representative of the single-stage target detection paradigm, has attracted academic attention due to its high-speed detection and real-time response capabilities. For example, Li et al. [20] introduced the exponential moving average (EMA) attention mechanism and the BiFPN-P feature fusion module based on YOLOv8, achieving a mean average precision of 91.5% in insulator positioning and small target defect detection, with a detection rate of 113 FPS.

Ahmed et al. [13] adopted a two-stage fine-tuning strategy. First, they used 18,000 images of 12 types of insulators to build a basic dataset to train the SSD model, and then optimize it through special scene datasets, such as different lighting and background conditions, effectively alleviating the problems of data scarcity and category imbalance. However, this method still has limitations in its ability to detect sub-pixel microcracks. Zhu et al. [21] proposed a dual variable convolutional kernel attention channel (DVCAC) in the YOLOv8 framework. By dynamically adjusting the convolution kernel scale and attention weight, they significantly enhanced the model's feature capture capability for multi-scale targets. Wei et al. [22] integrated the CBAM channel-spatial dual attention mechanism based on YOLOv5 and combined it with GIoU loss function to solve the loss calculation problem when the predicted box and the true box have no intersection, thereby improving the insulator defect detection accuracy by 5% and the detection speed to 142 FPS. Das and Leung [23] combined non-uniform attention mechanism, large receptive field and boundary refinement unit, and designed a loss function based on focal Tversky function, which improved the learning ability of small targets such as complex crack features and achieved 85% precision and 83% recall.

3. Materials and Methods

3.1. Overall Research Framework

As shown in Figure 1, we designed a phased research plan to meet the core requirements of small target detection and model lightweighting in insulator defect detection. First, we collected insulator-related images using a DJI M300 UAV and annotated the dataset with the LabelImg tool. Second, to preserve small-target features and achieve efficient feature fusion, we designed a Focus module to realize early aggregation of shal-

low and deep features. Meanwhile, we developed a new connection method based on two Focus modules to construct an Aggregation Diffusion Neck, ultimately obtaining the teacher model. Next, we compressed the teacher model via Group-Level First-Order Taylor Expansion Importance Assessment to generate the student model. To compensate for the performance loss caused by model compression, we further designed a Channel Distribution Distillation strategy. During this process, we determined the optimal design scheme for each component through ablation experiments, model pruning comparison experiments, and knowledge distillation comparison experiments. Finally, FocusNet was constructed, and its superiority in insulator defect detection tasks was verified through comparative experiments with mainstream models such as YOLOv5n, YOLOv8n, YOLOv11n, YOLOv12n, RT-DETR, RT-DETR-resnet18, RT-DETR-resnet34, Swin-Transformer-YOLO, RT-DETR-CSwinTransformer.

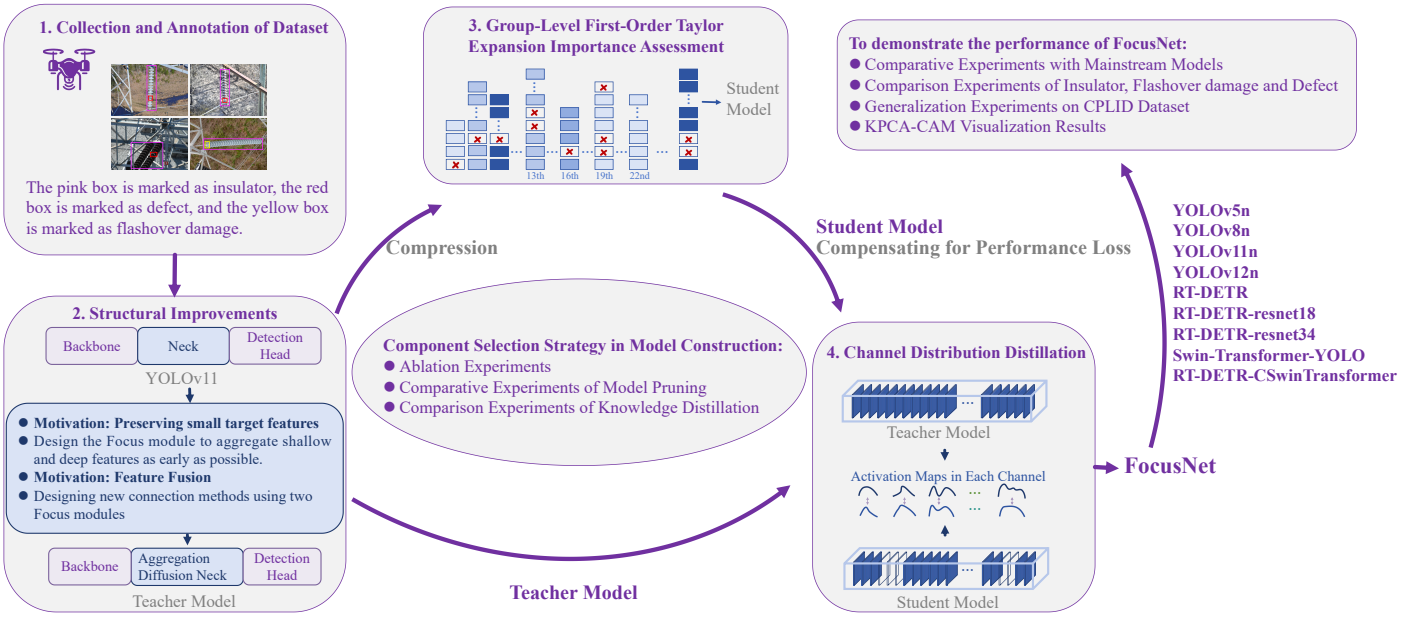


Figure 1. Research Roadmap. The framework follows a phased process: 1. Dataset construction (UAV image collection with DJI M300 + LabelImg annotation); 2. Teacher model development (Focus module design + Aggregation Diffusion Neck construction); 3. Model compression (Group-Level First-Order Taylor Expansion pruning to generate student model, where the red "X" symbol indicates that the channel in the convolution layer is removed during the pruning process); 4. Performance compensation (Channel Distribution Distillation strategy); 5. Scheme optimization (ablation, pruning, and distillation comparison experiments); 6. Model verification (comparative experiments with mainstream models like YOLO series and RT-DETR series).

3.2. Insulator Dataset

We used a DJI M300 UAV for aerial photography with the anti-shake function enabled. This data collection method provides a wide coverage of inspection areas and a high-altitude perspective. When constructing the insulator defect dataset, we focused on defect type characteristics, environment and background features, and shooting angle and distance characteristics to ensure its applicability to actual inspection scenarios. Specifically, for 110 kV overhead transmission lines, the safe flight distance of the UAV was set to 5–7 m, which not only ensures a safe distance from live equipment but also allows images to clearly capture insulator details. The shooting angle was mainly downward looking, with an angle of 30°–80° relative to the horizontal line, covering key areas of the top and sides of the insulator to avoid missing defects from a single perspective. In terms of environmental collection, we collected images across different seasons, time periods (morning, noon, and afternoon), and weather conditions (sunny, cloudy, and overcast) to simulate light

variations in real-world inspections. For background coverage, we included both natural scenes (rivers, vegetation, mountains, and farmland) and power-related man-made scenes (transmission tower steel structures, and conductors) to replicate complex inspection environments and enhance the model's anti-interference capability. After acquiring the insulator images, we annotated them using LabelImg. The dataset was divided into a training set, a validation set, and a test set, containing 1316, 264, and 372 images each, respectively. Detailed information is shown in Table 1. In addition, data augmentation techniques were applied during model training to further improve the robustness of the algorithm.

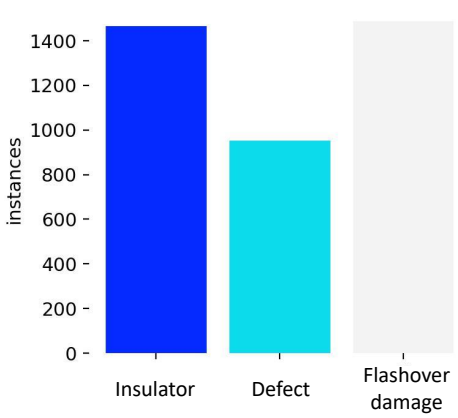
Table 1. Insulator dataset.

Category	Insulator	Defect	Flashover Damage
Instance number	2183	1460	2146
Image pixels	2144 × 1424	4298 × 3264	2216 × 2136
Shooting distance	5–7 m (consistent for all categories)		
Shooting angle	30°–80° (Angle from horizontal plane)		
Shooting background	Rivers, vegetation, mountains, farmland, transmission tower, conductors		
Weather conditions	Sunny, cloudy, overcast		

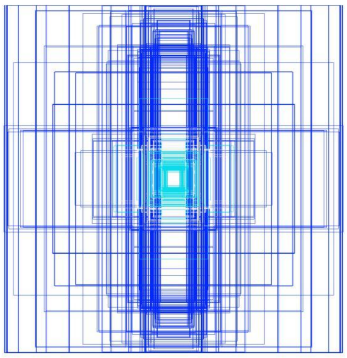
Figure 2 shows the visualization results of the insulator dataset. Figure 2a shows the number of instances of the three categories in the dataset. Specifically, the number of Insulator and Flashover damage instances exceeds 1400, and the number of Defect instances is close to 1000. Figure 2b shows schematic diagrams of the bounding boxes for the three categories. It can be observed that, on average, the bounding boxes of Defect and Flashover damage are smaller in size compared to those of Insulator: most cyan-colored bounding boxes (representing Defect) and light gray ones (representing Flashover damage) appear smaller, while blue ones (representing Insulator) are generally larger or more elongated. Figure 2c visualizes the distribution of target locations (with coordinates normalized to the range [0, 1]); it shows a higher concentration of targets in the central region (around $x \in [0.4, 0.6]$, $y \in [0.4, 0.6]$) compared to the edges, indicating that targets tend to cluster more in the center of the image. Figure 2d shows the normalized width and height distribution of objects. The x-axis represents width, and the y-axis represents height, both ranging from 0 to 1. Data points are primarily distributed in the ranges of $width \in [0, 0.2]$ and $height \in [0, 0.2]$, showing a distribution trend of small widths and small heights. This indicates that most objects in the dataset are small in size. Consequently, the model would learn these size characteristics during training, making it more sensitive to objects in this size range during detection.

3.3. YOLOv11 Algorithm

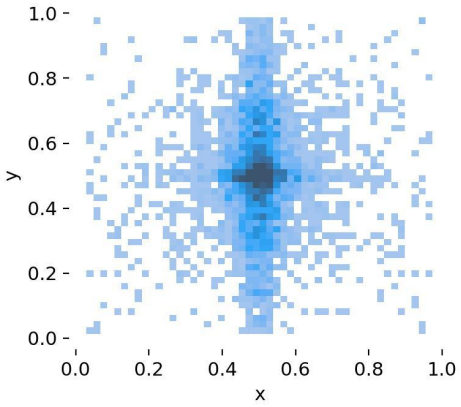
YOLOv11 is an object detection algorithm that can simultaneously predict the probability of the category and locate the bounding box. In the backbone network, the C3k2 module is configured with two smaller convolutional units instead of a single large convolutional unit, which improves computational efficiency while ensuring stable performance. At the same time, the C2PSA block is introduced to enhance the spatial attention mechanism, so that the model can focus more effectively on important areas in the image. In the detection head, YOLOv11n uses multiple C3k2 blocks to process multi-scale features. These C3k2 blocks are distributed in different paths, and the structure can be flexible according to the c3k parameters. The CBS layer is used to further refine the feature map to provide better feature input for the final detection task. Therefore, we designed a new framework for insulator defect detection based on YOLOv11n.



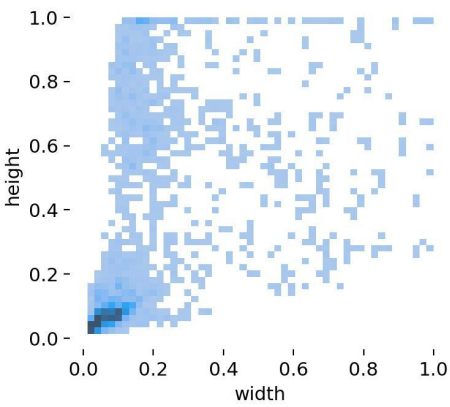
(a) Category Histogram. The horizontal axis represents the category, and the vertical axis represents the number of labeled instances within that category. Different categories are indicated by different colors: Insulator is blue, Defect is cyan, and Flashover damage is light gray.



(b) Visualization Example of Bounding Boxes. The first 500 bounding boxes in the dataset are selected, and the centers of all boxes are uniformly aligned to the center of a 1000×1000 pixel white canvas. The outline color of each box matches the color of the corresponding object category in (a).



(c) Scatter Plot of Bounding Box Center Locations. The horizontal axis represents the normalized x-coordinate of the bounding box center, and the vertical axis represents the normalized y-coordinate of the bounding box center. Both coordinates are normalized based on the image size and have units ranging from 0 to 1. The normalized coordinates in the heatmap show the distribution density of all bounding box centers, with darker colors indicating a greater number of objects at that location.



(d) Scatter Plot of Bounding Box Width-Height Ratio Distribution. The width and height axes represent the ratio of the object box width and height relative to the image width and height. Both values are normalized to the image size and range from 0 to 1. The normalized values in the heatmap show the density of bounding boxes for different width-height combinations, with darker colors indicating a higher concentration of objects with a specific aspect ratio.

Figure 2. Labeling and Data Distribution. This figure presents a visual analysis of the insulator defect dataset, including the category count distribution, annotation box distribution, target centroid distribution, and width–height distribution of the detected targets.

3.4. Aggregation Diffusion Neck

When detecting insulator defects and flashovers that occupy a small pixel area, the original feature fusion mechanism of YOLOv11 struggles to fully mine and integrate multi-scale feature information, and failing to effectively capture the subtle features of small targets, affecting the detection effect. Therefore, we improved the YOLOv11 algorithm, proposed an improved solution based on the Aggregation Diffusion Neck, and introduced the Focus module, as shown in Figure 3. This module focuses on the feature information of small targets and highlights the feature expression of small targets by screening and fusing feature maps of different scales. Meanwhile, we optimized the feature pyramid network, adjusted the upsampling and downsampling processes, and improved the propagation and utilization efficiency of small target features.

As shown in Figure 4, x_1, x_2, and x_3 come from different stages of the backbone network and have different resolutions and semantic strengths. The Upsample Module upsamples the shallow feature x_1 by two times to increase the spatial resolution, and then reduces the dimension through 1×1 convolution. The middle-level features maintain

the original dimension through 1×1 convolution identity mapping. ADown adaptively downsamples and reduces the dimension of the deep feature x_3:

$$x_1' = \text{Conv}_1(\text{Upsample}(x_1)) \tag{1}$$

$$x_2' = \text{Conv}_2(x_2) \tag{2}$$

$$x_3' = \text{ADown}(x_3) \tag{3}$$

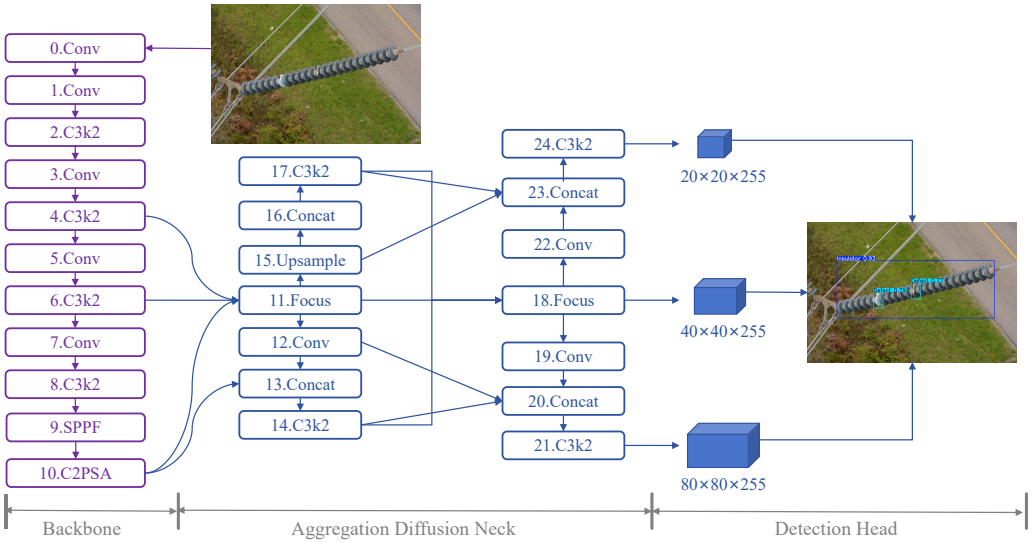


Figure 3. The structure of FocusNet. Aggregation Diffusion Neck is an innovation in the network structure. The Focus module and Concat aggregate feature maps with different numbers of channels and different receptive field sizes. The C3k2, Conv, and Upsample modules diverge feature maps to multiple aggregation blocks.

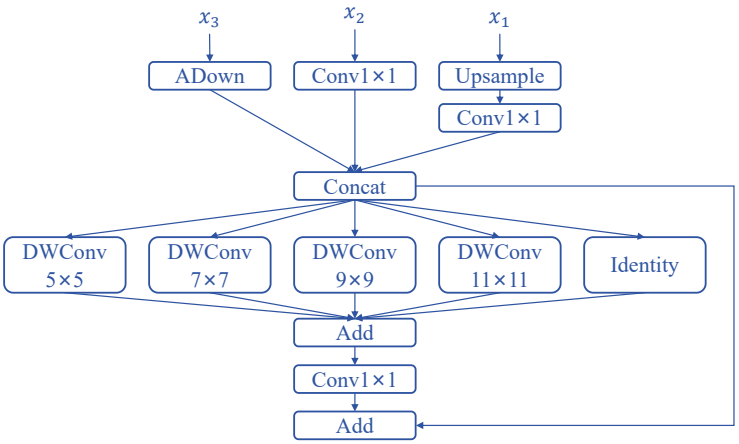


Figure 4. Focus. This is a neural network module structure for feature fusion and multi-scale feature extraction. ADown is adaptive downsampling, which is used to adjust the size of feature maps. 1×1 convolution is used to adjust the number of channels. Upsample is used to increase the resolution so that high-level features can be aligned with low-level features in size. DWConv contains four convolution kernels of different sizes and is used to extract features with different receptive fields. Identity represents directly passing the original features to retain basic information. Add means performing element-wise addition on the output tensors of multiple branches to fuse features with different receptive fields.

Use convolution kernels of different sizes (5×5, 7×7, 9×9, 11×11) to process the concatenated features in parallel to capture multi-scale local dependencies:

$$\text{Feature} = \text{Conv}_{1\times1}\left(\sum_{i=0}^{n}\text{DWConv}_{k_i}(x'_1 \oplus x'_2 \oplus x'_3)\right) \tag{4}$$

where $\oplus$ represents channel dimension concatenation, and DWConv_{k_i} represents depthwise separable convolution with kernel size k_i.

The residual connection adds the original concatenated features to the enhanced features to ensure that information is not lost and to stabilize the training:

$$\text{Output} = (x'_1 \oplus x'_2 \oplus x'_3) + \text{Feature} \tag{5}$$

In FocusNet, the output feature maps of two C3k2 modules and one C2PSA module in the backbone are all input into the Focus module of the Aggregation Diffusion Neck. The core advantage of this design lies in its ability to strengthen the correlation between features while improving the fusion efficiency of multi-scale information. The fine-grained detailed information extracted by C3k2 and the semantic information extracted by C2PSA can establish a strong correlation at the initial stage of feature fusion. The core challenge of small object detection is the easy loss of fine-grained features, especially during the multi-scale downsampling process. As a relatively shallow module in the backbone, C3k2 outputs high-resolution feature maps (which contain abundant detailed information about small objects). Forcing the binding and fusion of C3k2's high-resolution features and C2PSA's semantic features prevents fine-grained information from being excessively suppressed by downsampling operations in independent transmission paths.

3.5. Group-Level First-Order Taylor Expansion Importance Assessment Method

After the network structure was improved, the amount of calculation and the number of parameters increased slightly. The network actually used for insulator detection needs to consider reducing the model complexity. We proposed a group-level first-order Taylor expansion importance evaluation method to quantify the importance of pruning neural network neurons. The main goal is to minimize the error E:

$$\min_{W} E(\mathcal{D}, W) \tag{6}$$

where $W = w_0, w_1, \ldots, w_M$ are the neural network parameters, and the dataset $D = (x_0, y_0), (x_1, y_1), \ldots, (x_K, y_K)$ consists of input x_i and output y_i pairs.

Quantifying the importance of a parameter can be calculated by removing the error caused by it. The error can then be measured as the squared difference in prediction error with and without the parameter w_m:

$$\mathcal{I}_m = \left(E(\mathcal{D}, W) - E(\mathcal{D}, W \mid w_m = 0)\right)^2 \tag{7}$$

The importance estimation formula using the first-order Taylor expansion is

$$\mathcal{I}_m^{(1)} = \left(\frac{\partial E}{\partial w_m} \cdot w_m\right)^2 = (g_m \cdot w_m)^2 \tag{8}$$

where g_m is the gradient of the parameter w_m with respect to the loss function.

When estimating the joint importance of a set of structural parameters w_s, only considering the product of a single parameter and the gradient, the univariate approximation is

$$\widehat{\mathcal{I}}_{\mathcal{S}}^{(1)}(W) \triangleq \sum_{s\in\mathcal{S}} \mathcal{I}_s^{(1)}(W) = \sum_{s\in\mathcal{S}} (g_s w_s)^2 \tag{9}$$

The iterative pruning and fine-tuning process is as follows. First, gradient calculation and importance evaluation: for each mini-batch data, calculate the parameter gradient, and calculate the importance score of the neuron based on the gradient and weight. Second, importance aggregation and pruning: Accumulate the importance scores of multiple mini-batches, such as exponential moving average, momentum coefficient 0.9, and remove the N neurons with the lowest scores. Third, model fine-tuning: Continue training after pruning and repeat the above steps until the GFLOPs reach 50% of the original model. Figure 5 shows the pruning of each convolutional layer of FocusNet. Light purple represents the number of channels of the original convolutional layer, and dark purple represents the number of channels of the pruned convolutional layer. Most of the convolutional layers of FocusNet were pruned, with those in the FocusNet detection head undergoing extensive pruning.

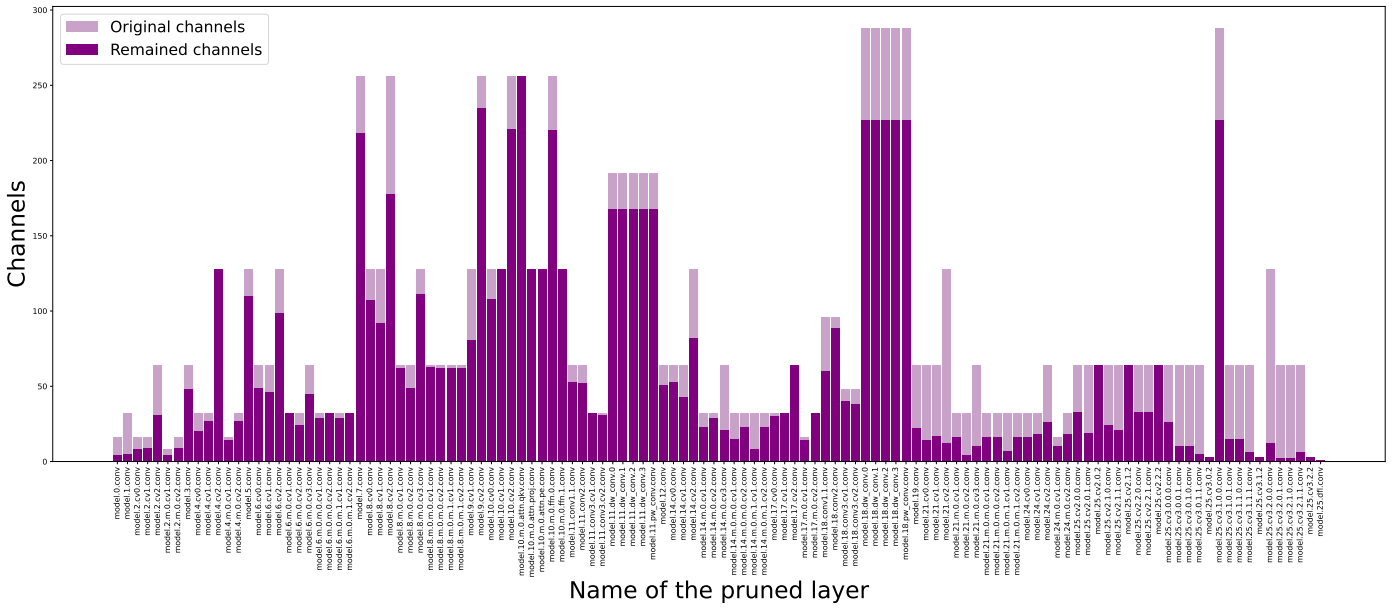


Figure 5. Group-Level First-Order Taylor Expansion Importance Assessment Method. The horizontal axis represents the name of the pruned layer, and the vertical axis represents the number of channels, ranging from 0 to 300. Two colors are used in the figure to distinguish data: light purple represents the number of channels in the original model, and dark purple represents the number of channels in the pruned model, visually showing the difference in the number of channels in each layer before and after pruning.

This pruning method explicitly quantifies the importance of a neuron as the squared change in loss after removing that neuron, which is expressed as $\mathcal{I}_m = (E(\mathcal{D}, W) - E(\mathcal{D}, W|w_m = 0))^2$. This definition directly links the impact of pruning on model performance, rather than making indirect inferences through weight magnitude, thus fundamentally solving the problem of disconnection between weight size and true importance. Meanwhile, the pruning algorithm features a unified evaluation scale across all layers, eliminating the need for per-layer sensitivity analysis. All network layers, including convolutional layers, linear layers, and skip layers in residual connections, share the same importance calculation standard. Even for the convolutional layers and skip layers within the residual blocks of the model, their importance evaluation requires no parameter adjustment, and high correlation is still maintained after pruning. In addition, the pruning algorithm incurs extremely low computational and memory overhead, as it only utilizes the gradient information already available during training without the need for additional forward or backward propagation operations.

3.6. Channel Distribution Distillation

The teacher model T was constructed based on the YOLOv11 framework by introducing the innovative Aggregation Diffusion Neck. The student model S was also built on the same YOLOv11 framework, integrating the Aggregation Diffusion Neck and then performing model compression via the Group-Level First-Order Taylor Expansion Importance Assessment Method. We used Channel Distribution Distillation for knowledge transfer as shown in Figure 6.

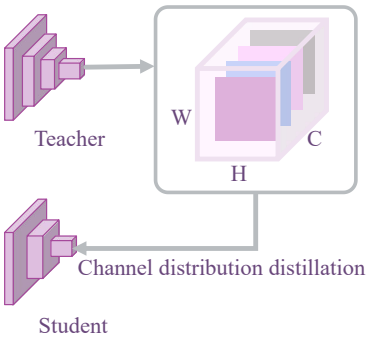


Figure 6. Knowledge Distillation. C represents the number of channels in the feature map, W represents the width of the feature map, and H represents the height of the feature map.

First, layer selection and feature extraction. We selected the 2nd, 4th, 6th, and 8th feature extraction layers (i.e., the four C3k2 layers) of the teacher model T and obtained their activation maps $y_c^T \in \mathbb{R}^{C\times W\times H}$. Meanwhile, we extracted the activation map y_c^S from the corresponding layers of the student model S, where $c = 1, 2, \ldots, C$ denotes the channel index.

Second, channel-level normalization and loss calculation. For each channel c of each layer, y_c^T and y_c^S were softmax normalized respectively to obtain probability distributions $\phi(y_c^T)$ and $\phi(y_c^S)$, where the temperature parameter T is usually set to 4. We calculated the Kullback–Leibler Divergence divergence loss for each layer:

$$\mathcal{L}_{\text{layer}} = \frac{\mathcal{T}^2}{C}\sum_{c=1}^{C}\sum_{i=1}^{W\cdot H}\boldsymbol{\phi}(\mathcal{G}_{c,i}^T)\cdot\log\left(\frac{\boldsymbol{\phi}(\mathcal{I}_{c,i}^T)}{\boldsymbol{\phi}(\mathcal{I}_{c,i}^S)}\right) \tag{10}$$

where $c = 1, 2, \ldots, C$ denotes the channel index, C represents the total number of channels in the feature map, and i indexes the spatial location of a channel. $\mathcal{T}$ stands for temperature, a hyperparameter that controls the smoothness of the probability distribution: the higher the temperature, the smoother the distribution, enabling each channel to focus on a wider spatial area. This loss forces the channel distribution of the corresponding layer of the student model to align with the teacher model, especially the semantically salient areas with high activation values.

Third, multi-layer loss aggregation and optimization. The weighted sum of the distillation losses of the 2nd, 4th, 6th, and 8th layers yielded a total distillation loss of

$$\mathcal{L}_{\text{distill}} = \sum_{k\in\{2,4,6,8\}}\alpha_k\cdot\mathcal{L}_{\text{layer}_k} \tag{11}$$

where α_k is the weight of the loss of each layer.

The total distillation loss was combined with the task loss of the student model (Focal Loss for Object Detection) to optimize the student model end-to-end:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{task}} + \mathcal{L}_{\text{distill}} \tag{12}$$

Channel Distribution Distillation changes the paradigm of knowledge distillation for small object prediction tasks from the spatial domain to the channel domain. Traditional spatial distillation methods must align all spatial locations, including redundant background regions, while channel-wise probability distributions can naturally highlight the high-activation regions of each channel, i.e., the pixels that are more critical for category prediction. The asymmetry of the KL divergence makes the loss calculation focus more on salient semantic regions with higher probabilities in the teacher network. Traditional symmetric losses force the student network to align with the teacher network in all regions. When the student network is small in scale, its capability is often insufficient to meet the requirements of such comprehensive alignment, and it is prone to performance degradation due to over-constraint. In contrast, the asymmetric KL divergence allows the student network to have a certain degree of deviation in non-critical regions and only needs to maintain alignment with the teacher network in core semantic regions, which is more in line with the learning law of small models gradually absorbing knowledge.

3.7. KPCA-CAM Visual Interpretability

To identify the key image regions on which the model's decisions depend, this study introduced Kernel PCA-based Class Activation Mapping (KPCA-CAM) to visualize the salient regions the model focuses on. KPCA-CAM constructs class activation mapping based on Kernel PCA. Its core approach is to map features into a high-dimensional space using a kernel function to capture nonlinear patterns. The specific steps are as follows.

Feature maps C were extracted from layers 11, 14, and 17 of FocusNet. The features of these layers contain multi-scale semantic information about the target, such as small object details and outlines of medium and large objects, which is key to the model's positioning and classification.

Then, the pairwise similarity of the eigenvectors in each layer of the feature map C was calculated to generate the kernel matrix K. This study used Radial Basis Function (RBF) kernel, which is defined as

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right) \tag{13}$$

where x_i, x_j is the eigenvector in the feature map, $\|\cdot\|^2$ is the squared Euclidean distance, and γ is the kernel parameter (set to 0.001) to control the sensitivity to local features.

Eigenvalue decomposition was performed on the kernel matrix K:

$$K = V \Lambda V^{-1} \tag{14}$$

where V is an orthogonal eigenvector matrix and Λ is a diagonal eigenvalue matrix.

The first eigenvector V_1 (corresponding to the largest eigenvalue) was selected, and the kernel matrix was projected onto V_1 to generate the class activation activation map L:

$$L = KV_1 \tag{15}$$

The kernel matrix K encodes the similarity between local image regions and the model's learned defect features (e.g., flashover edges and texture anomalies in insulators). Thus, projecting K onto V_1 (the eigenvector that captures the most variance in K) converts this similarity information into the intensity of map L. Specifically, the intensity of the map L directly reflects the contribution of the corresponding image region to the decision making of the model: since a higher intensity in L indicates a stronger similarity between the local region and the learned defect features, a stronger map L further indicates a more critical region for defect identification. In insulator flashover detection, the map L intensity in the

flashover region is significantly higher than the background, intuitively demonstrating the model's reliance on defect characteristics.

In FocusNet, layers 11, 14, and 17 process features of different resolutions (from high to low), respectively. Applying the above steps to each layer independently generates multi-scale activation maps that fully reflect the model's attention to target features at different levels.

3.8. Experimental Design, Experimental Environment and Evaluation Indicators

We selected 16 comparative models, and their selection reasons are explained in detail. Classic object detection models include SSD, YOLOv5n, YOLOv8n, YOLOv11n, YOLOv12n, and Improved Faster R-CNN [19]. These models are representative of traditional CNN-based detection architectures (single-stage for the YOLO series/SSD, two-stage for Faster R-CNN) and are widely used as baselines in object detection. Comparing with them can verify whether our method has advantages over mature classic frameworks. Mainstream industrial hybrid models focus on the RT-DETR series, including RT-DETR (original version) and RT-DETR with ResNet backbones of different depths (RT-DETR-resnet18, RT-DETR-resnet34) [17]. Selecting this series can confirm whether our method is competitive with current industrial-grade efficient detection models. Domain-specific improved models include Improved YOLOv3 [24], Improved YOLOv4 [25], ID-YOLO [26], and Insu-YOLO [27]. These models are specially optimized for insulator/defect detection and are widely cited in existing related studies. Comparing with them can directly reflect the advancement of our method in the specific field of power equipment detection. To further explore the advantages of Transformer architecture in power equipment detection, we also constructed two types of new comparative detectors: Swin Transformer [28] was embedded into the YOLOv11 framework as the backbone network to obtain Swin-Transformer-YOLO, and CSwin Transformer [29] was embedded into the RT-DETR framework as the backbone network to obtain RT-DETR-CSwinTransformer.

The L1 loss distillation and the L2 loss distillation are common knowledge distillation methods. The L1 loss is Mean Absolute Error (MAE), which calculates the average of the absolute values of the output differences between the student model and the teacher model. The formula is $\mathrm{L1}(s,t) = \frac{1}{N}\sum_{i=1}^{N}|s_i - t_i|$, where s is the output of the student model, t is the output of the teacher model, and N is the output dimension. The L2 loss is Mean Squared Error (MSE), which calculates the average of the squares of the output differences between the student model and the teacher model. The formula is $\mathrm{L2}(s,t) = \frac{1}{N}\sum_{i=1}^{N}(s_i - t_i)^2$.

All experiments were conducted on a workstation equipped with an NVIDIA GeForce RTX 3090 (24 GB video memory). The software environment was configured as follows: Ubuntu 18.04 operating system, CUDA 11.7 (for GPU acceleration), Python 3.8.16, and PyTorch 1.13.1 (Deep Learning Framework). This setup ensures the efficiency and stability of model training and inference. The key hyperparameters used in the training phase are summarized in Table 2.

Table 2. Hyperparameter settings for model training.

Hyperparameters	Value	Hyperparameters	Value
Image size	640 × 640	Epoch	500
Weight decay	0.0005	Batch size	32
Momentum	0.937	Learning rate	0.01

To comprehensively evaluate the performance of the insulator defect detection task, we adopted four widely used metrics in object detection: Precision (P%), Recall (R%), mAP@0.5 (mean Average Precision at an IoU threshold of 0.5), and mAP@0.5:0.95 (mean Average

Precision across IoU thresholds from 0.5 to 0.95 with a step of 0.05). The mathematical definitions of these metrics are provided below.

Precision (P%) quantifies the accuracy of positive predictions, i.e., the proportion of correctly predicted positive samples (insulator defects) among all predicted positive samples. It is calculated as follows:

$$P\% = \left(\frac{TP}{TP + FP} \right) \times 100\% \tag{16}$$

where TP (True Positives) denotes the number of insulator defect samples correctly classified as positive; FP (False Positives) denotes the number of non-defect samples incorrectly classified as positive.

Recall (R%) measures the completeness of positive detection, i.e., the proportion of correctly detected positive samples among all actual positive samples in the dataset. It is calculated as follows:

$$R\% = \left(\frac{TP}{TP + FN} \right) \times 100\% \tag{17}$$

where FN (False Negative) denotes the number of actual insulator defect samples incorrectly classified as negative.

Average Precision (AP) integrates Precision and Recall across all possible decision thresholds, reflecting the overall performance of the model for a single class. It is defined as the area under the Precision–Recall (P-R) curve:

$$AP = \int_0^1 P(R), dR \tag{18}$$

where $P(R)$ represents the Precision value corresponding to a given Recall value R, and the integral computes the area under the continuous P-R curve.

The mean Average Precision (mAP) extends AP to multi-class scenarios (e.g., different types of insulator defects: cracks and flashover damage). It is the average of AP values across all target classes:

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \tag{19}$$

where N denotes the number of target defect classes, and AP_i denotes the AP value of the i-th defect class. For mAP@0.5 and mAP@0.5:0.95, the above AP calculation is performed under the corresponding IoU thresholds, and then averaged to obtain the final mAP values.

4. Results

4.1. Comparative Experiments with Mainstream Models

As shown in Table 3, the precision of FocusNet was higher than YOLOv8n and RT-DETR-resnet18, indicating that it had the lowest false alarm rate . Traditional models (SSD and Faster R-CNN) lagged behind modern models in terms of accuracy. FocusNet's recall was 98.5%, ahead of the second-place RT-DETR-CSwinTransformer (98.2%). In terms of the indicator mAP @ 0.5, FocusNet ranked first with a score of 99.2%, higher than RT-DETR (98.9%) and YOLOv8n (98.8%), indicating that its complete target detection precision was the best under the medium threshold (IoU = 0.5). It was precisely because Aggregation Diffusion Neck solved the problems of a high small target missed detection rate caused by scale changes and complex background in insulator detection, thereby improving the overall robustness of the model. The computational load of FocusNet was the lowest among all the compared models. Due to the collaborative optimization strategy of channel pruning and knowledge distillation that we proposed, the model achieved extreme lightweight

while retaining strong feature expression ability and robustness. As shown in Figure 7, the background information in the sample dataset is complex, and the pixels of defects and flashovers occupy a relatively small proportion of the image area. Figure 8 shows the visual comparison of detection results among RT-DETR, YOLOv11, YOLOv12, and FocusNet in complex backgrounds such as forests, complex lines, grass, and tennis courts. In the forest background, FocusNet achieved a confidence score of 0.84 for detecting flashover damage, which was higher than that of other models. In the tennis court background, FocusNet achieved a confidence score of 0.99 for detecting insulators, which was also higher than that of other models. These results demonstrated that FocusNet has advantages in addressing issues such as missed detection of small object defects and interference from complex backgrounds.

Table 3. Comparative experiments of mainstream models. Black bold font indicates the optimal result, and "–" means missing data.

Model	P%	R%	mAP@0.5%	mAP@0.5:0.95%	Params	GFLOPs	Memory
SSD	89.28	73.78	84.11	–	34.31	–	–
Faster R-CNN	58.61	88.75	81.43	–	41.53	–	–
YOLOv5n	97.00	94.80	97.90	78.00	2.20	5.80	4.70
YOLOv8n	97.30	97.30	98.80	81.10	2.70	6.80	5.70
YOLOv11n	95.60	96.90	98.50	80.50	2.60	6.30	5.50
YOLOv12n	97.30	95.10	98.60	79.70	2.60	6.30	5.50
RT-DETR	96.70	95.70	98.90	–	2.50	–	–
RT-DETR-resnet18	98.40	97.50	98.40	80.20	19.90	56.90	80.70
RT-DETR-resnet34	97.90	97.90	98.90	79.00	31.10	88.80	125.70
Swin-Transformer-YOLO	96.90	96.60	98.30	78.40	29.70	77.60	60.00
RT-DETR-CSwinTransformer	97.20	98.20	99.20	80.80	30.50	89.90	123.20
FocusNet	**98.50**	**98.50**	**99.20**	**83.50**	**1.40**	**3.80**	**3.40**

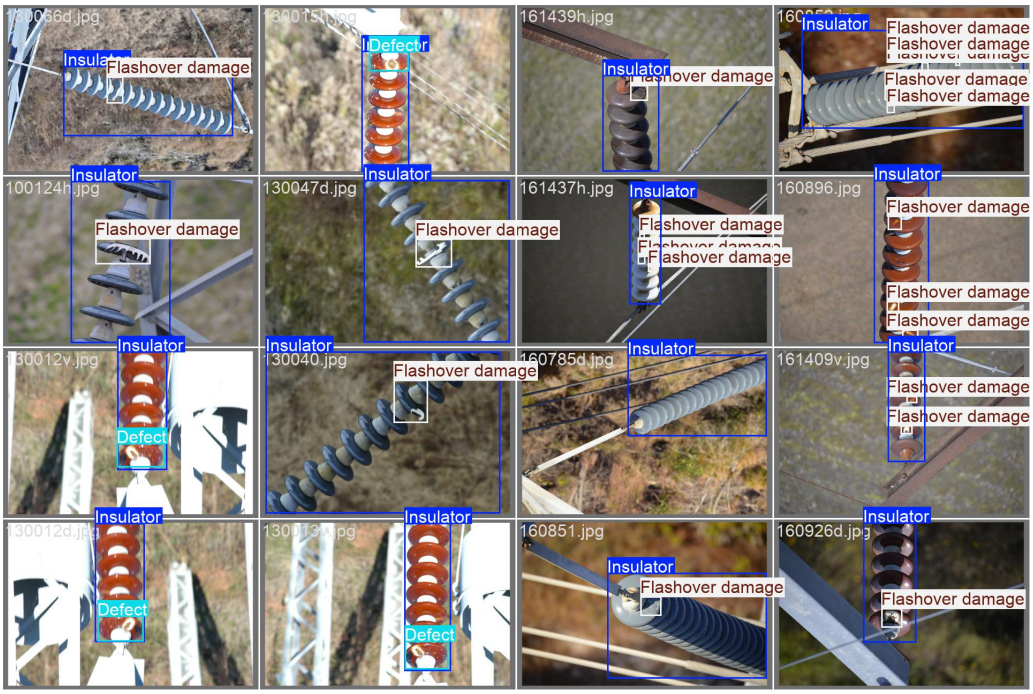


Figure 7. Verification Demonstration. These are visualizations of batch results from the validation phase of our self-built insulator dataset. The images show insulators of varying colors and structures, with the locations of the insulators and defects clearly circled in each image. Flashover damage, a frequently occurring defect, is labeled white in many images. Defects, which occur less frequently, are labeled cyan.

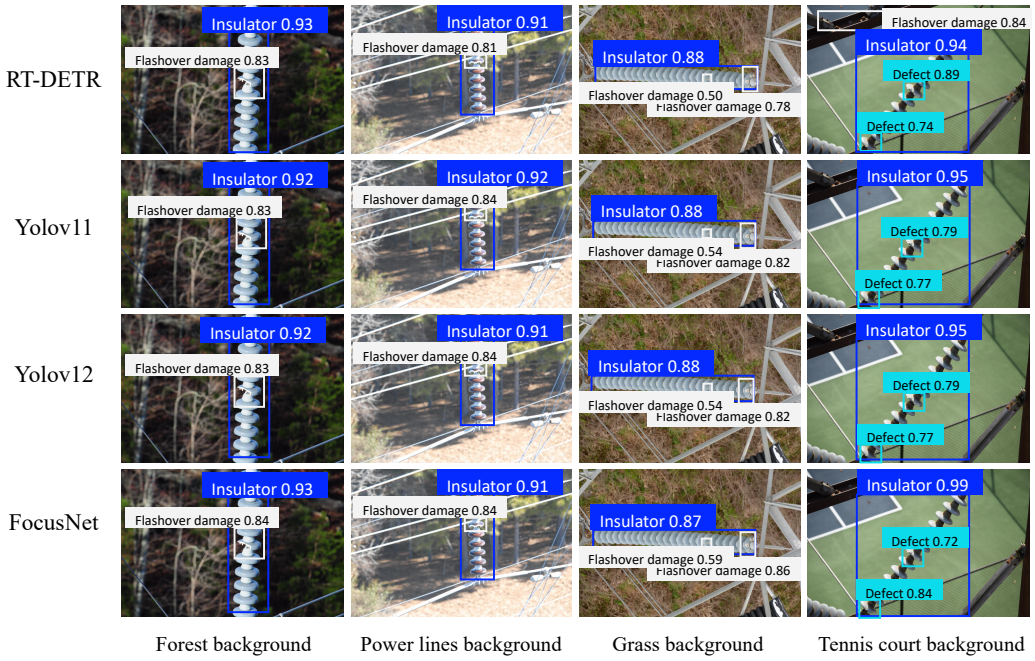


Figure 8. Detection Results. These are the detection results of different object detection models for insulator, flashover damage, and defect in four different backgrounds (forest, power lines, grassland, and tennis court). The results include the category names and confidence levels. Insulators are marked in blue, defects in cyan, and flashover damage in white.

4.2. Ablation Experiments

As shown in Table 4, three core components—Aggregation Diffusion Neck, Group Taylor (Group-Level First-Order Taylor Expansion Importance Assessment Method), and CDD (Channel Distribution Distillation)—were gradually introduced into the baseline model (YOLOv11n) to verify component compatibility and collaborative improvement effects. Aggregation Diffusion Neck increased the model's recall from 96.9% to 97.8% (+0.9) and mAP@0.5:0.95 from 80.5% to 82.7% (+2.2). It enhanced the model's stability across different scenarios and IoU thresholds, verifying the effectiveness of techniques, including early detail-semantic binding, multi-scale key feature transmission, and secondary aggregation purification. Group Taylor reduced the model parameters from 2.7 MB to 1.4 MB and GFLOPs from 7.6 to 3.8, verifying the effectiveness of combining iterative pruning and fine-tuning strategies. Channel Distribution Distillation improved recall from 98.2% to 98.5% and mAP@0.5:0.95 from 82.1% to 83.5%. It retained the lightweight advantages of the pruned architecture, confirming the effectiveness of Channel Distribution Distillation in balancing performance and efficiency. As shown in Figure 9, after adding three innovative points, the detection precision reached the optimal level after 500 rounds of training.

Table 4. Ablation experiment.

Model	R%	mAP@0.5:0.95%	Params	GFLOPs
YOLOv11n	96.90	80.50	2.60	6.30
YOLOv11n+ Aggregation Diffusion Neck	97.80	82.70	2.70	7.60
YOLOv11n+ Aggregation Diffusion Neck+ Group Taylor	98.20	82.10	1.40	3.80
YOLOv11n+ Aggregation Diffusion Neck+ Group Taylor+ CDD	98.50	83.50	1.40	3.80

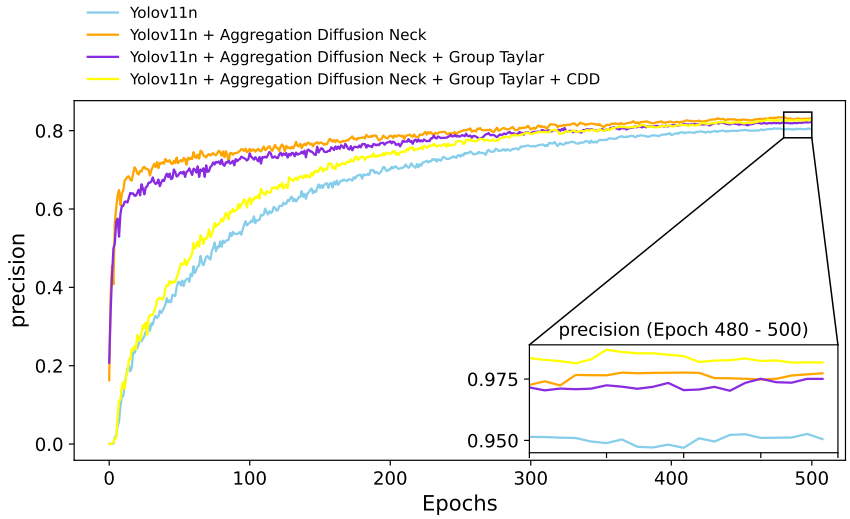


Figure 9. Precision Changes during training. This figure shows the changing trend of the precision of different improved versions of the Yolov11n model during training with the number of training epochs. At the same time, the zoomed-in figure focuses on the accuracy details in the later stages of training (Epoch 480–500).

4.3. Comparison Experiments of Model Pruning

As shown in Table 5, to verify the advancedness of our pruning method, the proposed Group Taylor was compared with three existing mainstream pruning methods, namely Slim [30], Group Norm [31], and Group Slim [31]. Group Taylor exhibited the best overall performance, with the highest recall (98.5%), mAP@0.5 on par with Slim (99.2%), and the number of parameters and memory usage comparable to Group SL (1.4 MB and 3.4 MB respectively). In contrast, Group Norm and Slim tended to be highly robust and extremely lightweight, respectively. Experimental results demonstrated that Group Taylor achieves effective lightweightness while maintaining core detection performance, making it the most practical choice among the four pruning methods.

Table 5. Comparative experiments of model pruning. Black bold font indicates the optimal result.

Model	R%	mAP@0.5%	Params	Memory
Slim [30]	97.60	99.10	1.40	**3.10**
Group norm [31]	96.50	98.30	2.00	4.40
Group sl [31]	98.10	98.10	1.50	3.40
Group Taylor	**98.50**	**99.20**	**1.40**	3.40

4.4. Comparison Experiments of Knowledge Distillation

As shown in Table 6, CDD was compared with existing distillation methods (MIMIC [32], L1 loss distillation, L2 loss distillation). CDD exhibited the best overall performance, ranking first in precision (98.5%), mAP@0.5 (99.2%), and mAP@0.5:0.95 (83.5%), and its recall (98.5%) was only 0.2 percentage points away from the best the best-performing L1 loss distillation (98.7%). Combined with the premise that all methods have the same efficiency, CDD maximized the learning of knowledge from the teacher model without increasing resource consumption. These results confirmed that CDD is the best choice among the four distillation methods.

Table 6. Comparative experiments on knowledge distillation. Black bold font indicates the optimal result.

Model	P%	R%	mAP@0.5%	mAP@0.5:0.95%
MIMIC [32]	97.30	98.40	99.10	83.30
L1	97.20	**98.70**	98.90	83.40
L2	97.30	98.3	98.80	82.90
CDD	**98.50**	98.50	**99.20**	**83.50**

4.5. Comparison Experiments of Insulator, Flashover Damage and Defect

Table 7 shows the performance of different models in three types of detection tasks: insulator, flashover damage, and defect. Insulator was detected with best precision, recall, mAP@0.5, and mAP@0.5:0.95 by FocusNet. This result demonstrated that FocusNet's anti-interference ability in complex backgrounds was significantly superior to that of models in the YOLO series and RT-DETR series. Flashover damage is a typical small target with blurry edges and low contrast against the insulator background. Flashover damage was detected with best precision, mAP@0.5, and mAP@0.5:0.95 by FocusNet. This result demonstrated that Aggregation Diffusion Neck enables small flashover target features to undergo a process of initial locking (first Focus fusion), enhanced expression (small-scale feature generation after the first fusion), directional transfer (secondary sub-scale path), and interference-free splicing (same-scale concatenation)—ensuring that small target features are not overwhelmed during the fusion stage. Defect has a large scale variation and complex texture features. Defect was detected by FocusNet, which achieved a precision of 98.70%—ranking first. This precision was 0.20 percentage points higher than that of RT-DETR (98.50%) and 1.20–3.10 percentage points higher than those of the YOLO series models (97.50% for both YOLOv5n and YOLOv8n, and 95.60% for YOLOv11n). In conclusion, through the collaborative optimization of the Aggregation Diffusion Neck, Group Taylor, and Channel Distribution Distillation, FocusNet met the engineering requirements of low misjudgment and high-precision localization for small targets.

Table 7. Comparison experiments of insulator, flashover damage, and defect.

	Model	P%	R%	mAP@0.5%	mAP@0.5:0.95%
Insulator	YOLOv5n	98.60	99.30	99.20	94.80
	YOLOv8n	98.50	99.70	99.40	95.70
	YOLOv11n	98.30	99.70	99.50	96.00
	RT-DETR	98.60	99.70	99.50	90.30
	RT-DETR-resnet18	98.60	100.00	99.50	89.70
	FocusNet	**98.90**	**100.00**	**99.50**	**96.40**
Flashover damage	YOLOv5n	95.00	90.20	95.70	66.80
	YOLOv8n	96.00	90.20	97.50	67.80
	YOLOv11n	93.00	93.60	97.10	69.10
	RT-DETR	95.40	98.80	98.40	73.40
	RT-DETR-resnet18	97.60	**99.60**	98.40	73.20
	FocusNet	**97.90**	96.90	**98.50**	**73.70**
Defect	YOLOv5n	97.50	94.70	98.90	72.40
	YOLOv8n	97.50	95.40	98.80	75.60
	YOLOv11n	95.60	97.40	98.90	75.90
	RT-DETR	98.50	98.50	**99.50**	77.90
	RT-DETR-resnet18	97.40	99.00	99.40	77.90
	FocusNet	**98.70**	**99.40**	99.40	**80.30**

4.6. Generalization Experiments on CPLID Dataset

To verify FocusNet's adaptability to external data, we conducted generalization experiments on the CPLID dataset ([33]) as shown in Table 8. The selected comparative models include YOLOv8n, YOLOv11n, RT-DETR, RT-DETR-resnet18, as well as published improved models in this field (Improved Faster R-CNN [19], Improved YOLOv3 [24], Improved YOLOv4 [25], ID-YOLO [26], Insu-YOLO [27]), so as to ensure the comprehensiveness and representativeness of the comparison. FocusNet achieved 98.9% precision, a 0.9 percentage point improvement over Improved YOLOv3 (98.00%) and 0.9 percentage point higher than YOLOv12n (98.00%), which achieved outstanding accuracy. FocusNet achieved 99.2% recall, significantly outperforming Improved YOLOv3 (95.00%) and Improved YOLOv4 (93.99%), and surpassing mainstream lightweight models such as YOLOv5n (97.90%), YOLOv8n (97.40%) and RT-DETR (98.10%). FocusNet also achieved an mAP@0.5 of 99.20%, tying with RT-DETR (99.20%) for the best result, and surpassed traditional improved models such as Improved YOLOv4 (97.26%), Improved YOLOv3 (96.50%) and Insu-YOLO (95.90%). mAP@0.5 comprehensively reflects the detection stability of a model at different confidence thresholds. This result demonstrated FocusNet's outstanding stability and accuracy in defect and flashover damage identification. FocusNet's parameter count was only 1.40 M, fewer than ID-YOLO (5.90 M). FocusNet's computational complexity (GFLOPs) was 3.80, significantly lower than Insu-YOLO (13.80 GFLOPs) and all Transformer-based models (such as RT-DETR: 103.4 GFLOPs, RT-DETR-resnet18: 56.90 GFLOPs, and Swin-Transformer-YOLO: 77.60 GFLOPs), making it the most computationally efficient of all compared models.

Table 8. Generalization experiments on CPLID dataset. Black bold font indicates the optimal result, and "–" means missing data.

Model	P%	R%	mAP@0.5%	Params	GFLOPs
Improved Faster R-CNN [19]	–	–	92.00	–	24.00
Improved YOLOv3 [24]	98.00	95.00	96.50	–	–
Improved YOLOv4 [25]	93.80	93.99	97.26	–	–
ID-YOLO [26]	92.14	–	95.60	5.90	–
Insu-YOLO [27]	–	–	95.90	–	13.80
YOLOv5n	96.40	97.90	99.10	2.10	5.80
YOLOv8n	95.40	97.40	99.00	2.60	6.80
YOLOv11n	95.10	98.40	98.90	2.50	6.30
YOLOv12n	98.00	96.30	99.00	2.56	6.30
RT-DETR	96.40	98.10	99.20	32.00	103.4
RT-DETR-resnet18	94.70	97.40	99.10	19.87	56.90
RT-DETR-resnet34	96.80	97.20	99.00	31.11	88.80
Swin-Transformer-YOLO	97.20	98.40	99.10	29.72	77.60
RT-DETR-CSwinTransformer	95.90	97.20	98.90	30.49	89.90
FocusNet	**98.90**	**99.20**	**99.20**	**1.40**	**3.80**

4.7. KPCA-CAM Visualization Results

KPCA-CAM [34] generates class activation maps by projecting the kernel matrix of the convolutional layer feature maps onto the first principal component, thus mapping high-dimensional nonlinear features back to the original image space to visualize the attention distribution of the model during detection. Figure 10 presents the experimental results based on this visualization technique. To verify the superiority of the aggregated diffusion neck in learning the defect features of transmission line insulators, we conducted tests on YOLOv8, YOLOv11, YOLOv12, and FocusNet. The results showed that the high-response regions in FocusNet's activation maps highly overlap with the core discriminative features of the target. Specifically, for small-scale defects and flashover targets, FocusNet's activation maps maintained precise focus without being diluted by background information. This was

consistent with the advantage of KPCA-CAM in capturing nonlinear spatial relationships in feature maps. FocusNet's performance confirmed that its perception mechanism for small-scale features and weight distribution mechanism are reasonable, enabling it to learn meaningful target features. This further verified that KPCA-CAM models nonlinear patterns in feature maps through kernel functions, providing reliable visual evidence for evaluating the model's ability to capture key target features.

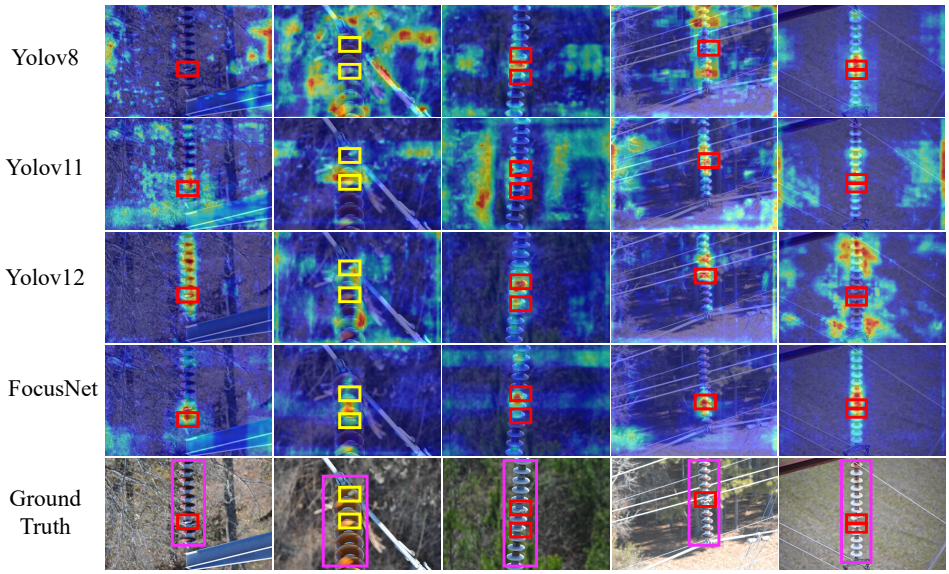


Figure 10. KPCA-CAM visual interpretability. The red areas in the heat map represent high-response areas that the model focuses on, and the blue areas are low-response background. The pink box is marked as insulator, the red box is marked as defect, and the yellow box is marked as flashover damage.

Furthermore, KPCA-CAM can be directly applied to CNN-based defect recognition models onboard drones, without relying on model gradients or modifying network architectures, generating explanatory maps in real time. By comparing the original image with the activation map, inspectors can verify the rationality of the model's predictions and enhance their confidence in the inspection results. Furthermore, if the model misjudges an object, the map can assist in analyzing the cause, such as misidentifying background interference as a defect, providing a basis for manual review and model optimization.

5. Discussion

5.1. Advantages and Limitations of the Proposed Model Components for Insulator Defect Detection

This study combined structural improvement, model pruning, and knowledge distillation methods to optimize the performance of deep learning models for insulator defect detection, achieving a new balance strategy between efficiency and accuracy. Aggregation Diffusion Neck ensures that the semantic information and spatial features of small objects are not overwhelmed by aggregating multi-scale features and diffusing contextual information. However, it increased computational effort by 20.6% and the number of parameters by 3.8%. Group Taylor estimates neuronal importance through first-order derivatives ($\mathcal{I}_m^{(1)} = (g_m w_m)^2$), removing redundant filters while preserving task-critical features. Unlike weight-based pruning or batch normalization-based scaling, Group Taylor adaptively identifies redundant neurons by quantifying their contribution to loss reduction, avoiding the oversimplification of simply equating importance with weight size. Although pruning reduced computational overhead, it could disrupt the learned feature hierarchy. Channel Distribution Distillation addresses this issue by transferring knowledge from a large teacher model to a pruned student model, leveraging the teacher's soft labels to opti-

mize the student's decision boundary. Compared to traditional compression techniques, distilling knowledge into the pruned student model reduces the computational cost of the distillation process itself. The performance of knowledge distillation depends heavily on the temperature parameter, which must be calibrated to balance the teacher's soft labels with the student's hard targets. An inappropriate temperature can preserve noise in the teacher's output.

5.2. Practical Deployment Feasibility on UAV Platforms

In terms of model lightweighting, FocusNet has a parameter count of only 1.40 M, a computational speed of 3.80 GFLOPs, and a memory footprint of 3.40 MB. Compared with RT-DETR-resnet18, which has 19.90 M parameters and a memory footprint of 56.90 GFLOPs, FocusNet shows a significant reduction in resource consumption. The edge computing unit mounted on the DJI M300 supports a maximum real-time inference capability of 8 GFLOPs, 8 GB of memory, and 64 GB of storage space, which fully meets the resource requirements of FocusNet. In terms of generalization, FocusNet achieved stable and high-performance detection results on both the self-built insulator defect dataset and the publicly available CPLID dataset. On the self-built dataset, it achieved precision of 98.50%, recall of 98.50%, and mAP@0.5 of 99.20%. On the CPLID dataset, it achieved mAP@0.5 of 99.20% (vs. 98.90% for YOLOv11n). These results demonstrate that FocusNet is not overfitted to specific scenarios and can adapt to the needs of insulator inspections in different regions and under different shooting conditions. The model is implemented based on PyTorch, which can be converted to ONNX format for deployment on mainstream edge frameworks (e.g., TensorRT, OpenVINO), supporting both GPU and CPU environments.

5.3. Challenges of Small Object Detection in Insulator Scenes

Water film or droplets formed by rain or dew can change the surface texture of the insulator and cause light scattering, reducing the contrast between small defects and the background. Wei et al. [22] addressed similar pain points by adopting GIoU loss and Mish activation, achieving a minimum small-defect AP of 37.5% in complex environments—but their model had a parameter count of 5.2 M (3.7× that of FocusNet) and a computational speed of 12.3 GFLOPs (3.2× that of FocusNet). In contrast, FocusNet maintains lightweight advantages (1.40 M parameters, 3.80 GFLOPs) but lacks a dedicated reflection/humidity noise suppression module, leading to a 5.3% drop in small-defect AP under strong sunlight (vs. a 2.1% drop for Wei's model) and a 7.1% drop in heavy rain (vs. a 3.8% drop for Wei's model). This highlights FocusNet's need for improved environmental robustness while preserving lightweight performance.

5.4. Future Work

In future work, we plan to develop dynamic pruning ratios based on layer sensitivity, such as retaining more filters in early convolutional layers, which may further improve efficiency. At the same time, we consider integrating the pruning criterion directly into the distillation loss function (such as penalizing low-importance neurons during the distillation process), which may simplify the process and reduce training time. In addition, we plan to add a lightweight Environmental Adaptation Attention Module. This module will learn to weight defect features differently under light reflection/wet surface conditions, and integrate it into the FocusNet Head layer to further improve small object detection robustness in complex environments.

6. Conclusions

For features like insulator defects and flashovers that occupy an extremely small proportion in images, traditional object detection algorithms have obvious limitations

in recognition capability. They tend to yield poor recognition results due to the low pixel proportion of the targets and weak feature information. This study proposes a new framework, FocusNet, based on YOLOv11, which improves this approach by combining feature enhancement with model lightweighting. We design an Aggregation Diffusion Neck and introduce a Focus module to strengthen the multi-scale feature fusion mechanism and enhance the representation of small object features. The Group-Level First-Order Taylor Expansion Importance Assessment Method is adopted to prune redundant channels that contribute little to detection accuracy, significantly compressing model parameters. Combined with Channel Distribution Distillation technology, the accuracy loss of the pruned model is controlled within a very small range. On a self-built insulator dataset, FocusNet's detection recall is increased by 1.6 percentage points, achieving a mAP@0.5 score of 99.20%, both significantly outperforming YOLOv11. Furthermore, FocusNet achieves 97.90% accuracy in flashover damage detection and 98.70% accuracy in defect detection. Compared to the YOLOv11 baseline, its computational complexity (3.80 GFLOPs) and memory usage (3.40 MB) are reduced by 40% and 38%, respectively, meeting the requirements for edge device deployment. This research provides an efficient and reliable technical solution for the intelligent monitoring of power equipment.

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

Funding: This work was supported by Basic Research Project of Liaoning Provincial Department of Education (LJ212410147042), Basic Research Projects of the Department of Science & Technology of Liaoning Province (2022JH2/101300274) and Research Project on Teaching Reform of Graduate Education in Liaoning Province (LNYJG2023117).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Acknowledgments: During the writing of this manuscript, the author(s) used Doubao to polish the initial sections of the introduction and discussion.

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

Abbreviations

The following abbreviations are used in this manuscript:

Group Taylor	Group-Level First-Order Taylor Expansion Importance Assessment Method
CDD	Channel Distribution Distillation

References

1. Li, W.; Hu, J.; Lu, G.; Huang, X. Topology optimization of acoustic bandgap crystals for topological insulators. *Eng. Comput.* **2024**, *40*, 2581–2594. [CrossRef]
2. Mitrovic, M.; Titov, D.; Volkhov, K.; Lukicheva, I.; Kudryavzev, A.; Vorobev, P.; Li, Q.; Terzija, V. Supervised learning based method for condition monitoring of overhead line insulators using leakage current measurement. *Eng. Appl. Artif. Intell.* **2025**, *143*, 110040. [CrossRef]
3. Gouda, O.E.; Darwish, M.M.F.; Mahmoud, K.; Lehtonen, M.; Elkhodragy, T.M. Pollution severity monitoring of high voltage transmission line insulators using wireless device based on leakage current bursts. *IEEE Access* **2022**, *10*, 53713–53723. [CrossRef]
4. Ilomuanya, C.; Nekahi, A.; Farokhi, S. A study of the cleansing effect of precipitation and wind on polluted outdoor high voltage glass cap and pin insulator. *IEEE Access* **2022**, *10*, 20669–20676. [CrossRef]

5. Xia, C.; Pan, F.; Hu, Q.; Long, D.; Rong, W.; Wei, Y.; Zhou, Z. Review of Insulator Wet Flashover Modelling and Factors Affecting Characteristics. In Proceedings of the 2023 IEEE Sustainable Power and Energy Conference (iSPEC), Chongqing, China, 28–30 November 2023; pp. 1–13.
6. Zhong, Y.; Hu, R.; Li, Z.; Cai, Y. Insulator Defect Detection Based on YOLOv4-tiny with Improved Feature Fusion. In Proceedings of the 2022 12th International Conference on Information Technology in Medicine and Education (ITME), Xiamen, China, 18–20 November 2022; pp. 273–277.
7. Zhang, Z.; Yang, S.; Jiang, X.; Ma, X.; Huang, H.; Pang, G.; Ji, Y.; Dong, K. Hot water deicing method for insulators Part 2: Analysis of ice melting process, deicing efficiency and safety distance. *IEEE Access* **2020**, *8*, 130729–130739. [CrossRef]
8. Zhang, L.; Wang, L.; Yan, Z.; Jia, Z.; Wang, H.; Tang, X. Star generative adversarial VGG network-based sample augmentation for insulator defect detection. *Int. J. Comput. Intell. Syst.* **2024**, *17*, 141. [CrossRef]
9. Liu, Y.; Huang, X.; Liu, D. Weather-domain transfer-based attention YOLO for multi-domain insulator defect detection and classification in UAV images. *Entropy* **2024**, *26*, 136. [CrossRef]
10. Chen, M.; Tian, Y.; Xing, S.; Li, Z.; Li, E.; Liang, Z.; Guo, R. Environment perception technologies for power transmission line inspection robots. *J. Sens.* **2021**, *2021*, 5559231. [CrossRef]
11. Xi, Y.; Zhou, K.; Meng, L.-W.; Chen, B.; Chen, H.-M.; Zhang, J.-Y. Transmission line insulator defect detection based on swin transformer and context. *Mach. Intell. Res.* **2023**, *20*, 729–740. [CrossRef]
12. Feng, F.; Yang, X.; Yang, R.; Yu, H.; Liao, F.; Shi, Q.; Zhu, F. An Insulator defect detection network combining bidirectional feature pyramid network and attention mechanism in unmanned aerial vehicle images. *Eng. Appl. Artif. Intell.* **2025**, *152*, 110745. [CrossRef]
13. Ahmed, M.F.; Mohanta, J.C.; Sanyal, A. Inspection and identification of transmission line insulator breakdown based on deep learning using aerial images. *Electr. Power Syst. Res.* **2022**, *211*, 108199. [CrossRef]
14. Liu, Y.; Liu, D.; Huang, X.; Li, C. Insulator defect detection with deep learning: A survey. *IET Gener. Transm. Distrib.* **2023**, *17*, 3541–3558. [CrossRef]
15. Li, X.; Blancaflor, E.B. A review of image-based insulator defect detection algorithms for transmission lines. In Proceedings of the 2024 5th International Conference on Computer Vision, Image and Deep Learning (CVIDL), Zhuhai, China, 19–21 April 2024; pp. 718–724.
16. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
17. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
18. Zhou, M.; Wang, J.; Li, B. ARG-mask RCNN: An infrared insulator fault-detection network based on improved mask RCNN. *Sensors* **2022**, *22*, 4720. [CrossRef] [PubMed]
19. Zhao, W.; Xu, M.; Cheng, X.; Zhao, Z. An insulator in transmission lines recognition and fault detection model based on improved faster RCNN. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 5016408. [CrossRef]
20. Li, Z.; Jiang, C.; Li, Z. An insulator location and defect detection method based on improved yolov8. *IEEE Access* **2024**. [CrossRef]
21. Zhu, C.; Cao, H.; Li, T.; Li, J.; Wang, Y.; Wang, S. Multi-scale self-attention network: A light and fast circuit detection method. *Electr. Power Syst. Res.* **2025**, *247*, 111716. [CrossRef]
22. Wei, L.; Jin, J.; Deng, K.; Liu, H. Insulator defect detection in transmission line based on an improved lightweight YOLOv5s algorithm. *Electr. Power Syst. Res.* **2024**, *233*, 110464. [CrossRef]
23. Das, A.K.; Leung, C.K.Y. A novel technique for high-efficiency characterization of complex cracks with visual artifacts. *Appl. Sci.* **2024**, *14*, 7194. [CrossRef]
24. Liu, J.; Liu, C.; Wu, Y.; Xu, H.; Sun, Z. An improved method based on deep learning for insulator fault detection in diverse aerial images. *Energies* **2021**, *14*, 4365. [CrossRef]
25. Qiu, Z.; Zhu, X.; Liao, C.; Shi, D.; Qu, W. Detection of transmission line insulator defects based on an improved lightweight YOLOv4 model. *Appl. Sci.* **2022**, *12*, 1207. [CrossRef]
26. Hao, K.; Chen, G.; Zhao, L.; Li, Z.; Liu, Y.; Wang, C. An insulator defect detection model in aerial images based on multiscale feature pyramid network. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3522412. [CrossRef]
27. Chen, Y.; Liu, H.; Chen, J.; Hu, J.; Zheng, E. Insu-YOLO: An insulator defect detection algorithm based on multiscale feature fusion. *Electronics* **2023**, *12*, 3210. [CrossRef]
28. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 10012–10022.
29. Dong, X.; Bao, J.; Chen, D.; Zhang, W.; Yu, N.; Yuan, L.; Chen, D.; Guo, B. Cswin transformer: A general vision transformer backbone with cross-shaped windows. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 12124–12134.

30. Liu, Z.; Li, J.; Shen, Z.; Huang, G.; Yan, S.; Zhang, C. Learning efficient convolutional networks through network slimming. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2736–2744.
31. Fang, G.; Ma, X.; Song, M.; Bi, M.M.; Wang, X. Depgraph: Towards any structural pruning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–27 June 2023; pp. 16091–16101.
32. Li, Q.; Jin, S.; Yan, J. Mimicking very efficient network for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6356–6364.
33. Tao, X.; Zhang, D.; Wang, Z.; Liu, X.; Zhang, H.; Xu, D. Detection of Power Line Insulator Defects Using Aerial Images Analyzed With Convolutional Neural Networks. *IEEE Trans. Syst. Man Cybern. Syst.* **2018**, *50*, 1486–1498. [CrossRef]
34. Karmani, S.; Sivakaran, T.; Prasad, G.; Ali, M.; Yang, W.; Tang, S. KPCA-CAM: Visual explainability of deep computer vision models using kernel PCA. In Proceedings of the 2024 IEEE 26th International Workshop on Multimedia Signal Processing (MMSP), West Lafayette, IN, USA, 2–4 October 2024; pp. 1–5.

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