Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería

Tech Science Press

https://doi.org/10.32604/rimni.2026.075481

ARTICLE

AMST-YOLO: An Improved YOLO11-Based Method for Insulator Defect Detection

Yanping Chen[1], Ran Guo[1], Le Zou[1], Junchao Liu[1], Zailu Shen[1] and Hailei Wang[2,*]

[1]School of Artificial Intelligence and Big Data, Hefei University, Hefei, 230601, China
[2]Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei, 230001, China

*Corresponding Author: Hailei Wang. Email: hlwang@iim.ac.cn

Received: 02 November 2025; Accepted: 19 December 2025; Published: 06 November 2026

ABSTRACT: For power systems to operate safely and steadily, it is essential to identify defects in insulators used on transmission lines. However, challenges such as strong background interference and significant scale variations in UAV aerial images often lead to suboptimal performance of existing detection methods, particularly for small-scale insulator defects. To address these issues, this paper proposes an improved YOLO11 detection method named AMST-YOLO. First, based on the YOLO11n algorithm, a lightweight convolutional module called ADown (Adaptive Downsampling) is introduced for downsampling in the network architecture. This adaptive downsampling approach reduces computational redundancy and enhances the model's sensitivity to small targets. Second, a Triplet Attention module is embedded at the end of the backbone network, which strengthens the network's capability to focus on salient information in complex backgrounds through an improved channel-weighting strategy. Third, a novel feature extraction module, CMSC (Contextual Multi-scale Spectral Convolution Module), is proposed to reduce feature redundancy while strengthening the model's multi-scale feature extraction capability. Finally, the Inner MPDIoU (Minimum Point Distance IoU) loss function is adopted to enhance the accuracy of bounding box regression, enabling precise localization and adaptation to small-scale targets of varying sizes. Experimental results demonstrate that the improved YOLO11 model achieves increases of 7%, 5.5%, and 5.2% in precision, recall, and mean average precision (mAP), respectively, compared to the baseline model. The proposed method effectively enables accurate identification of insulator defects.

KEYWORDS: Insulator defect detection; YOLO11; attention mechanism

1 Introduction

Insulators in transmission lines are critical components of power systems, serving the dual purposes of supporting and insulating transmission conductors to ensure the safe and stable delivery of electricity. However, during long-term operation, insulators are prone to defects such as cracks, stains, and breakages due to environmental pollution, mechanical damage, and high-temperature aging. These defects not only affect the operational efficiency of transmission lines but may also pose serious safety hazards [1]. Traditional inspection of transmission line insulators primarily relies on manual visual inspection by trained personnel, which is time-consuming, labor-intensive, and involves risks associated with working at heights [2–4]. With the widespread adoption of unmanned aerial vehicle (UAV) inspection technologies, the efficiency of image acquisition has been significantly improved. Nevertheless, aerial images are often plagued by severe background interference and large variations in target scales, which hinder accurate defect detection and

Copyright © 2027 The Authors. Published by Tech Science Press. This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

reduce detection reliability. Therefore, achieving efficient and precise detection of insulator defects is of great practical importance [5].

As deep learning continues to evolve, neural network–driven object detection has achieved remarkable advances in computer vision, thereby promoting progress in insulator defect detection. By learning features from large-scale data, it improves detection accuracy and robustness, and has been widely applied in insulator defect detection. Current object detection frameworks can generally be categorized into two-stage and one-stage approaches. Two-stage detection methods generally involve region proposal generation followed by classification and bounding box regression. Region Proposal Networks (RPNs) are first used to generate candidate regions, which are then classified and refined to achieve accurate object localization and identification. Owing to their high detection accuracy, two-stage methods have been widely used in insulator defect detection tasks. For example, existing studies have improved the Faster R-CNN architecture by integrating pixel-level segmentation networks, deploying the models on UAV platforms, or incorporating multi-scale feature fusion mechanisms [6–9] to address complex real-world scenarios.

In contrast, single-stage detection methods eliminate the need for region proposal generation and perform classification and regression directly within the network, significantly improving detection speed. This makes them better suited for real-time applications, particularly in UAV-based inspection scenarios. Representative single-stage algorithms include the YOLO (You Only Look Once) series [10], SSD (Single Shot MultiBox Detector) [11], and RetinaNet [12]. As one of the more advanced and widely adopted variants in the YOLO family, YOLO11 offers a strong balance among detection accuracy, inference speed, and model lightweighting. Its well-suited applications include defect detection tasks in UAV-based inspection scenarios. In insulator defect detection, YOLO11 demonstrates solid performance under relatively simple backgrounds and moderate object scales, achieving efficient and reliable detection. However, its performance degrades in the presence of complex backgrounds, dense noise, or small-scale insulator defects, leading to missed detections and false positives. Despite the improvements YOLO11 provides, detection networks based on YOLO11 still face several challenges. First, the downsampling strategy of its backbone causes fine-grained textures and subtle cracks—critical cues in insulator defect detection—to be prematurely lost, weakening its ability to detect small or weak-feature defects. Second, insulators are frequently located against complex backgrounds such as sky or metallic structures, which often exhibit textures highly similar to insulator surfaces. The existing attention mechanisms in YOLO11 cannot sufficiently highlight target regions under such background interference, resulting in misdetections or omissions. Third, the neck-stage feature fusion of YOLO11 may introduce redundant or inconsistent semantic features when dealing with slender, scale-varying insulator targets, limiting the effectiveness of its multi-scale representation. Finally, the standard bounding-box regression loss fails to fully capture the geometric constraints of small or elongated insulator components, reducing localization accuracy.

In response to these limitations, we develop a detection model that utilizes YOLO11n as its foundation. To reduce information loss and computational redundancy during downsampling—which often leads to small-object omission—a lightweight convolution module, ADown, is introduced for adaptive downsampling, enhancing small-target perception while improving efficiency. To strengthen key feature extraction in complex scenes—where insulators may share similar textures with the background or be partially obscured by noise—the Triplet Attention module is incorporated into the backbone to optimize channel weighting and reinforce attention to critical features. In addition, to reduce feature redundancy and enhance multi-scale fusion, a novel CMSC module is introduced in the feature fusion stage. This module integrates SCConv (Spatial and Channel reconstruction Convolution) and MSCA (Multi-Scale Convolutional Attention) to minimize redundancy and improve multi-scale feature representation. Finally, to enhance bounding-box

regression accuracy and cross-scale adaptability, the Inner MPDIoU loss is adopted, enabling more precise localization of small targets.

2 Related Work

Deep learning has become the cornerstone of modern insulator defect detection, inspiring numerous detection frameworks and enhancement strategies. Wang et al. [13] proposed the CSDN network model, which has been widely applied in high-resolution and complex-background insulator fault detection. Wu et al. [14] introduced a single-stage insulator instance defect segmentation method based on YOLACT (You Only Look At CoefficienTs), incorporating an attention mechanism and an improved feature fusion network to overcome the efficiency bottlenecks of segmentation algorithms.

Among various deep learning models, the YOLO series has attracted particular attention in insulator defect detection due to its excellent balance between speed and accuracy. Jiang et al. [15] proposed ACCP-YOLOv8 for insulator defect detection, introducing an additional small-object detection layer and enhancing shallow–deep feature fusion to improve detection accuracy for small defects. Li et al. [16] designed the LiteYOLO-ID algorithm, which employs a lightweight convolution module (EGC) and a lightweight backbone network to reduce parameters and accelerate feature fusion. Zhang et al. [17] proposed FINet for detecting insulators and defects generated by synthetic fog algorithms, introducing a channel attention mechanism into YOLOv5 to improve small-target defect detection accuracy under complex weather conditions. Cao et al. [18] enhanced the YOLOv8 model by integrating the Coordinate Attention Mechanism (CAM) and a Feature Channel Shuffle Operation (CSO) to address the loss of low-level details caused by rain and fog. Akindele et al. [19] proposed YOLO-ELA, which incorporates an Efficient Local Attention (ELA) block into the neck of the YOLOv8 architecture to increase the model's focus on defective insulators. Zhang et al. [20] combined a YOLOv5s-based detection network with a classification network composed of VGG16 and SPP-Net (Spatial Pyramid Pooling), achieving an insulator defect detection accuracy of 89%. To address the limitations of the original YOLO11 model—namely insufficient small-object detection capability and weak spatial feature extraction under complex image backgrounds—Ji et al. [21] proposed an improved YOLO11-based defect recognition algorithm for distribution-line insulators. Shen et al. [22] introduced an innovative Multi-Focused Feature Extraction Convolution (MFFEConv) module to replace the convolution blocks in the YOLO11 backbone, effectively improving detection performance under complex backgrounds, unstable lighting, and small defect sizes. Lan et al. [23] proposed RMC-YOLO, a YOLO11-based detection algorithm designed to overcome the limited small-object detection performance and insufficient multi-scale feature fusion of traditional methods. Liang et al. [24] incorporated a recurrent attention mechanism into YOLO11 to achieve accurate defect localization under real-time constraints on edge-deployment devices.

Although YOLO11 has demonstrated strong performance in insulator defect detection, its effectiveness remains limited under complex backgrounds and large scale variations, particularly for small or fine-grained defects. Feature loss during downsampling and insufficient discriminative capability further constrain its applicability in UAV-based inspection scenarios. To address these issues, this paper introduces the AMST-YOLO insulator defect detection method. By integrating adaptive downsampling, an attention mechanism, and enhanced multi-scale feature extraction, the proposed approach improves the accuracy of insulator defect detection to a considerable extent. This improved model is referred to as AMST-YOLO, where "A" stands for the ADown module, "MS" represents the multi-scale attention mechanism from the CMSC module, and "T" denotes the Triplet Attention structure.

3 Methodology

Although YOLO11 achieves satisfactory performance in many scenarios, challenges remain in complex backgrounds and small object detection, particularly in industrial applications such as insulator defect detection on power transmission lines, which demand higher recognition accuracy. To tackle these challenges, this paper proposes an enhanced YOLO11n-based method that improves the detection of small defects in complex environments. The structure of the AMST-YOLO network is illustrated in Fig. 1.

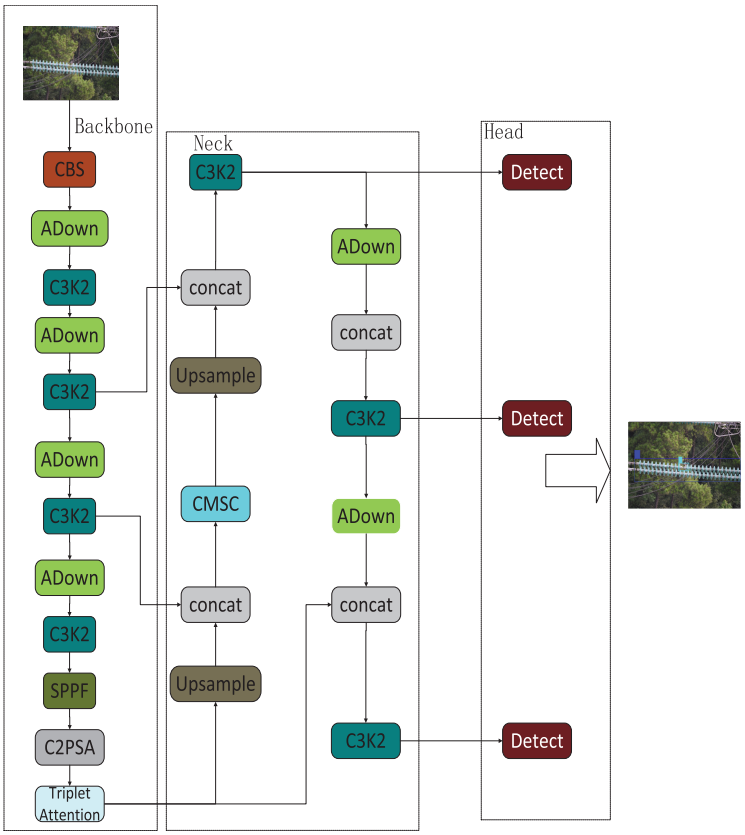


Figure 1: The structure of AMST-YOLO network.

First, a lightweight downsampling module, ADown, is employed in the backbone network. This module utilizes a lightweight convolutional structure combined with adjustable stride and channel compression mechanisms to effectively reduce computational redundancy during downsampling, while maximally preserving crucial feature information. This enhances the model's perception of small objects. Second, a Triplet Attention block is appended to the final stage of the backbone. By constructing three-dimensional cross-attention pathways across spatial, height, and channel dimensions, this mechanism improves the model's focus on key regions within complex backgrounds, thereby better distinguishing foreground targets from background interference. Third, a multi-scale feature extraction module named CMSC is designed by integrating SCConv and MSCA attention mechanisms. This module suppresses redundant features while enhancing feature discriminability, enabling the model to maintain strong detection performance even under scenarios with large variations in object scale. Finally, an improved bounding box regression loss function is deployed in the regression branch of the detection head, MPDIoU, which is based on the Inner-IoU concept. This loss function not only considers the spatial overlap between predicted and ground truth boxes but

also incorporates internal region overlap and scale adaptability of the target. As a result, the bounding box localization accuracy is significantly improved, particularly for small-sized objects.

3.1 ADown Downsampling Module

Conventional downsampling methods typically rely on strided convolutions or convolutional operations with large strides to reduce the resolution of feature maps. However, these approaches often result in significant loss of detail, especially when processing small objects. To address this issue, the improved YOLO11 network adopts the ADown [25] module for downsampling.

The ADown module enhances the richness of feature representation through feature partitioning and a multi-branch design, effectively reducing the loss of fine details. Its structure is illustrated in Fig. 2. Initially, a 2×2 average pooling operation with a stride of 1 and no padding is applied for smooth features and enhance local information. The resulting feature map is then split along the channel dimension into two parts. These two parts are processed separately through two branches. The first branch (X1) uses a 3×3 convolution with stride 2 to extract primary features and downsample the feature map. The second branch (X2) first applies a 3×3 max pooling (stride = 2) for downsampling and then a 1×1 convolution to adjust the number of channels. Finally, the outputs of both branches are concatenated along the channel dimension to form the output feature map. This multi-branch strategy effectively combines the advantages of convolution and pooling, preserving critical information while enhancing feature diversity.

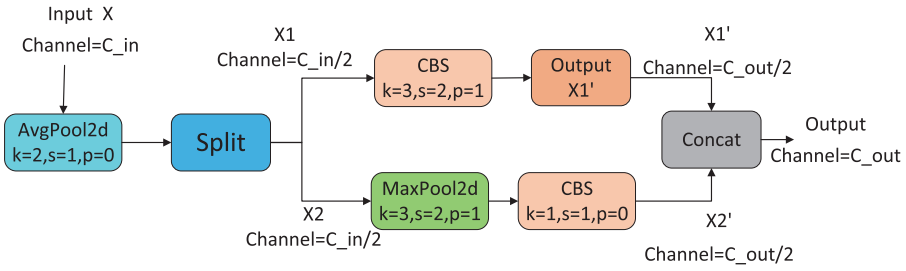


Figure 2: ADown structure.

3.2 Triplet Attention Module

In insulator defect detection for transmission lines, images often contain complex background interference such as vegetation, conductors, and towers. This complexity can easily lead to missed or false detections, especially when the defect targets are small or their features are not prominent. Traditional attention mechanisms, such as the Squeeze-and-Excitation (SE) module [26], enhance feature representation by modeling channel-wise dependencies. The Convolutional Block Attention Module (CBAM) [27] further improves focus on target regions by incorporating spatial attention. However, CBAM models channel and spatial attention separately, limiting its ability to capture multi-dimensional interactions, which reduces effectiveness under complex backgrounds. To enhance model robustness in such scenarios, this paper introduces the Triplet Attention (TA) module [28] at the end of the backbone network (as shown in Fig. 3). TA adopts a cross-dimensional interaction strategy to jointly model channel and spatial information, improving sensitivity to fine-grained defect features and small-object representation.

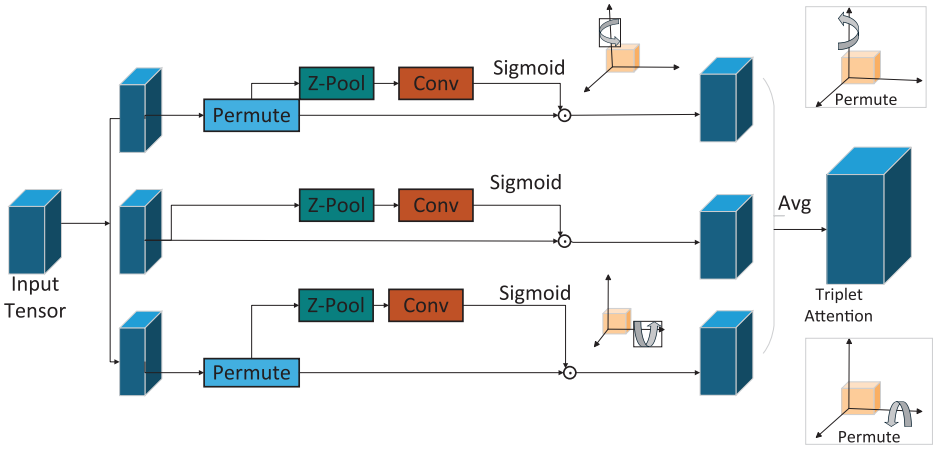


Figure 3: Triplet attention mechanism.

For a given input tensor $X \in \mathbb{R}^{C\times H\times W}$, the three different branches are used for processing: The first branch models the relationship between the channel dimension C and the spatial dimension W. To achieve this, the input tensor is first rotated 90° along the height axis (H-axis). Then, a Z-pool operation is applied to the rotated tensor, which performs both max pooling and average pooling simultaneously, followed by concatenation to obtain a 2D feature map:

$$\text{Z-pool}(X) = [\text{MaxPool}_{d=0}(X),\ \text{AvgPool}_{d=0}(X)] \tag{1}$$

where $\text{MaxPool}_{d=0}(\cdot)$ denotes max pooling along the 0-th dimension of the tensor, and $\text{AvgPool}_{d=0}(\cdot)$ denotes average pooling along the same dimension. Then, the result is passed through a convolution layer, normalization, and a Sigmoid activation function to generate the attention weights, which are applied to the rotated input tensor. Finally, the tensor is rotated back to its original dimensions. The second branch focuses on the relationship between height H and width W. It directly applies the Z-pool operation, followed by convolution and a Sigmoid activation function to generate spatial attention weights, which are then used to reweight the input tensor. The third branch aims to capture the relationship between the channel dimension C and the spatial dimension H. Its processing steps are similar to the first branch but use a different rotation direction and dimension-switching strategy. Finally, the outputs of the three branches are fused using an averaging strategy to obtain the final output tensor Y. The computation formula is as follows:

$$Y = \frac{1}{3}\left(\overline{\hat{x}_1 \cdot \sigma\left(\phi_1\left(\hat{x}_1^*\right)\right)} + \hat{x}_2 \cdot \sigma\left(\phi_2\left(\hat{x}_2\right)\right) + \overline{\hat{x}_3 \cdot \sigma\left(\phi_3\left(\hat{x}_3^*\right)\right)}\right) \tag{2}$$

where $X \in \mathbb{R}^{C\times H\times W}$ is the input tensor, $\hat{X}_i$ denotes the tensor after rotation in the i branch, and $\hat{X}_i^*$ represents the tensor after the Z-pool operation in the i branch. ϕ_1, ϕ_2, and ϕ_3 refer to standard 2D convolution layers in the three branches, defined by kernel size k, and σ denotes the Sigmoid activation function.

3.3 CMSC Module

In the inspection of transmission lines, due to the uncertainty in the shooting angle and altitude of UAVs, insulators and their defects often exhibit significant scale variations and sparse distributions. Although the C3K2 structure used in YOLO11 adopts a lightweight residual connection design that can reduce computational complexity and preserve semantic information to a certain extent, it still has two major limitations: First, its spatial modeling capability is limited, making it difficult to capture the spatial dependencies between complex objects; second, it lacks a multi-scale feature perception mechanism, which leads to performance degradation when dealing with objects of drastically varying sizes or small targets. To address these issues, this paper proposes an improved CMSC module. Based on retaining the backbone structure of the C3K2 module, the Bottleneck structure and output channel fusion strategy are optimized. Specifically, to overcome the weak spatial modeling capability and the high redundancy in the Bottleneck structure of the C3K2 module, the second standard convolution is replaced with an SCConv [29] module, forming a new SCB (SCConv and Convolution Block) module, as shown in Fig. 4.

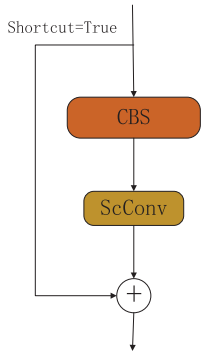


Figure 4: SCB module structure.

SCConv enhances spatial context modeling by enabling the network to more effectively perceive the spatial layout relationships between local and global objects, thereby improving its ability to recognize complex scene structures. To address the lack of scale-awareness in the C3K2 module, MSCA [30] module is introduced after concatenating the output channels of multiple SCB modules. The MSCA module adaptively assigns attention weights across multiple scales, reinforcing the model's responsiveness to small or blurry targets and significantly improving detection performance. The final CMSC structure, as illustrated in Fig. 5, retains the backbone framework of the C3K2 module while substantially enhancing spatial awareness and multi-scale adaptability.

3.3.1 SCConv Module

SCConv consists of a Spatial Reconstruction Unit (SRU) and a Channel Reconstruction Unit (CRU), aiming to reconstruct spatial and channel information within the feature maps to suppress redundant features and enhance the model's representational and multi-scale perception capabilities. The network structure of SCConv is illustrated in Fig. 6.

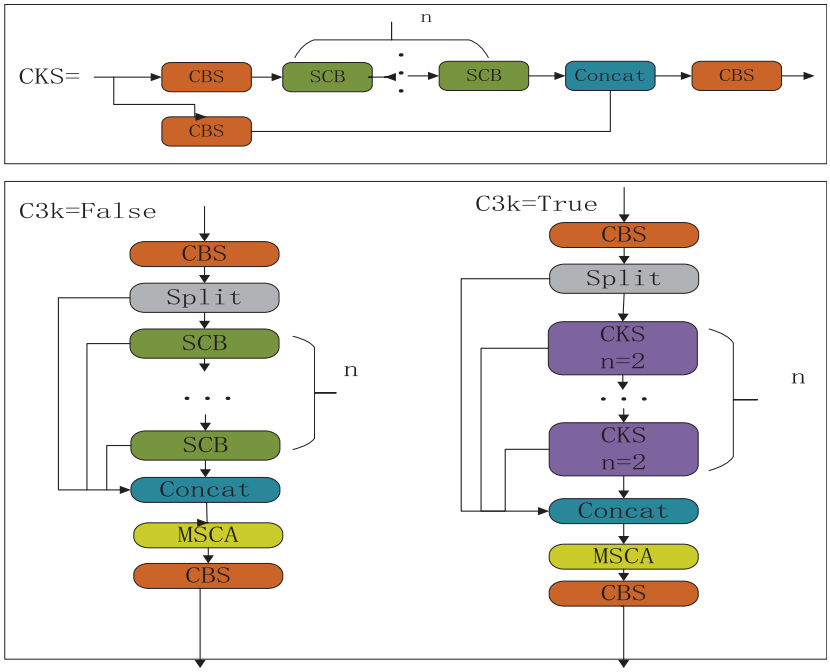


Figure 5: Network structure of CMSC.

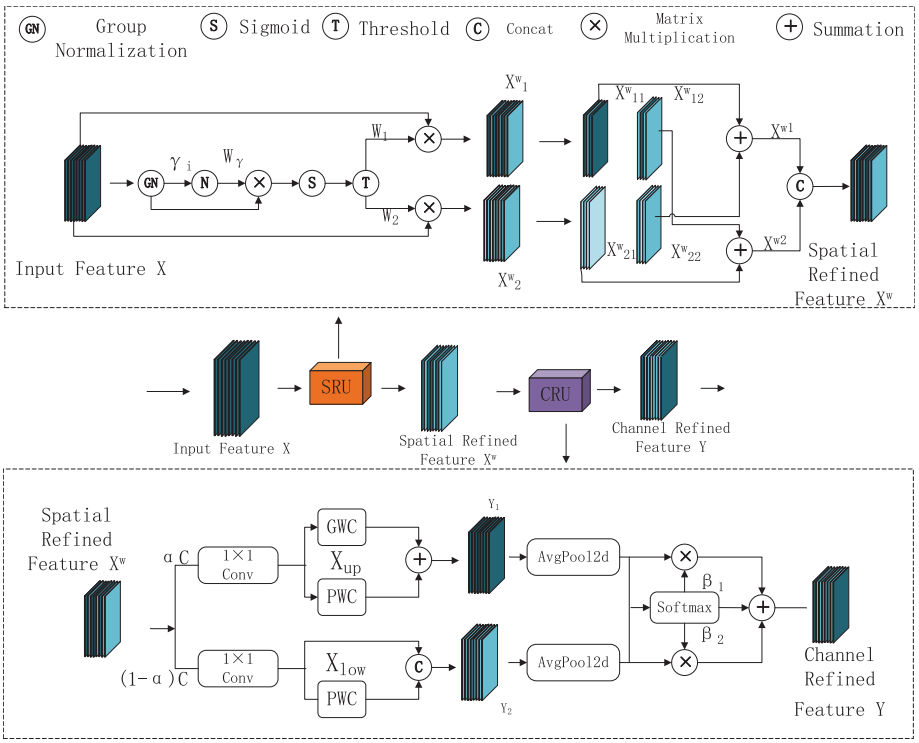


Figure 6: Network structure of SCConv.

The SRU module adaptively adjusts the channel weights of the input feature map to selectively enhance important features. Given an input feature map $X \in \mathbb{R}^{C\times H\times W}$ (where B is the batch size, C is the number of

channels, and H and W are the height and width of the feature map, respectively), SRU first applies Group Normalization to X. This normalization standardizes features within each channel group, mitigating scale discrepancies among features, enhancing training stability, and accelerating convergence. The computation is as follows:

$$\mathrm{GN}(X) = \frac{x - \mu}{\sqrt{\sigma^2 + \varepsilon}} \cdot \gamma + \beta \tag{3}$$

where μ and σ represent the mean and standard deviation, respectively, γ and β are the learnable parameters for each group, ϵ is a small constant to prevent division by zero. Furthermore, Formula (4) normalizes the learnable parameters to the range (0, 1) to obtain normalized correlation weights W_γ, which are used to measure the importance of different feature maps.

$$W_\gamma = \{w_i\} = \frac{\gamma_i}{\sum_{j=1}^{C} \gamma_j}, \quad i, j = 1, 2, \ldots, C \tag{4}$$

The normalized weights are multiplied with the group-normalized results and passed through a sigmoid activation function to constrain the output between 0 and 1. Next, the SRU applies a gating mechanism for selective weighting of the feature maps: when the weight exceeds a preset gate threshold of 0.5, it is set to 1; if it is below the threshold, it is set to 0. Through this gating mechanism, two sets of weights are obtained—W_1 with richer information content and W_2 with less information. The computation formulas are as follows:

$$W = \mathrm{Gate}(\mathrm{Sigmoid}(W_\gamma(\mathrm{GN}(X)))) \tag{5}$$

where Sigmoid is the activation function, and Gate refers to the threshold control performed by the gating mechanism. Using weights W_1 and W_2, the SRU divides the input feature map into two parts, respectively, weighted to obtain a feature map rich in information X_1^W and another with less information X_2^W. Finally, the weighted feature maps are fused through a feature reconstruction operation.

The CRU module mainly consists of three steps: channel partitioning, feature transformation, and feature fusion. In the channel partitioning stage, the input feature map $X \in \mathbb{R}^{C \times H \times W}$ is split along the channel dimension into two parts, with αC and $(1 - \alpha)C$ channels, respectively (where $\alpha \in (0, 1)$ is a tunable hyperparameter). Then, 1×1 convolutions are applied to reduce the number of channels, producing the upper branch feature X_{up} and lower branch feature X_{low}, which are used for extracting different types of information in subsequent stages. During the feature transformation stage, X_{up} is fed into a parallel structure composed of Group-Wise Convolution (GWC) and Point-Wise Convolution (PWC). The outputs of these two convolutions are summed to form the feature map Y_1. Meanwhile, the lower branch feature X_{low} passes through a PWC and is concatenated with itself to generate Y_2, which contains shallow details to supplement low-level information. In the fusion stage, global average pooling is performed on both feature groups to obtain channel descriptor vectors S_1 and S_2. Then, a SoftMax operation generates channel attention weights β_1 and β_2. These weights are used to reweight Y_1 and Y_2, which are then element-wise summed to output the final feature map Y.

3.3.2 MSCA Module

To further enhance the model's ability to perceive multi-scale objects, this paper introduces the MSCA attention mechanism within the CMSC module, which adaptively adjusts channel weights after integrating the output features from multiple SCB modules. MSCA effectively captures and utilizes multiscale information in the input features, enabling the model to achieve higher perception accuracy and discriminative capability when handling objects of varying sizes and complexities. Its structure, shown in Fig. 7.

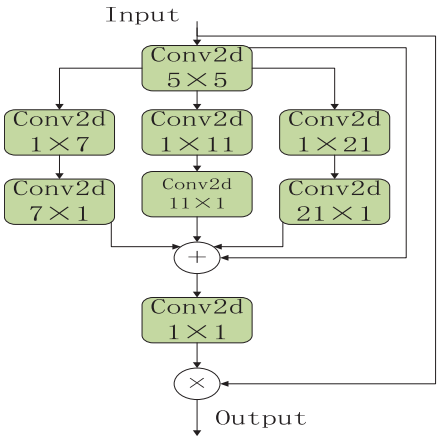


Figure 7: MSCA module structure.

First, the fused feature map passes through a 5 × 5 depthwise separable convolution to preliminarily capture spatial contextual information and enhance connections within local regions. Then, MSCA employs three groups of strip convolution paths with different scales, simulating receptive fields of 7 × 7, 11 × 11, and 21 × 21, respectively, to extract feature information at multiple spatial scales. Specifically, these convolution operations are decomposed into directional convolutions (e.g., $1 \times k$ and $k \times 1$), which reduce computational cost while maintaining an effective receptive field. Finally, the feature maps from the different scale paths are fused channel-wise and further modeled for inter-channel dependencies via a 1 × 1 convolution. The fused result is passed through a sigmoid activation function to generate a channel attention weight map, which is then element-wise multiplied with the original input features to enhance semantically significant regions and suppress redundant background information.

3.4 Optimization Loss Function Inner MPDIoU

In object detection tasks, Intersection over Union (IoU) is commonly used to measure the overlap between the predicted and ground truth bounding boxes. YOLO11 adopts Complete IoU (CIoU), which integrates IoU, the distance between center points, and the aspect ratio to improve modeling for targets of varying sizes and shapes. However, when the predicted and ground truth boxes have similar aspect ratios but different sizes, CIoU shows limited sensitivity to positional differences, resulting in weak regression signals. To address this, Ma et al. [31] proposed MPDIoU loss, which minimizes the distances between the top-left and bottom-right corners of the two boxes for more accurate discrepancy measurement. Compared to traditional IoU-based losses, MPDIoU incorporates overlap, center distance, and size differences, making it more sensitive to variations in target scale. Its calculation formula is as follows:

$$d_1^2 = (x_1^{\text{prd}} - x_1^{g_t})^2 + (y_1^{\text{prd}} - y_1^{g_t})^2 \tag{6}$$

$$d_2^2 = (x_2^{\text{prd}} - x_2^{\text{gt}})^2 + (y_2^{\text{prd}} - y_2^{\text{gt}})^2 \tag{7}$$

$$MPDIoU = IoU - \frac{d_1^2}{w^2 + h^2} - \frac{d_2^2}{w^2 + h^2}, \tag{8}$$

where (x_1, y_1) and (x_2, y_2) denote the top-left and bottom-right coordinates of the predicted box and ground truth box, respectively, w and h represent the width and height of the target, d_1, d_2 are the Euclidean distances between the corner points. Although MPDIoU performs well in bounding box regression, it lacks effective

perception of the internal structural distribution, especially during high-IoU stages where the loss changes minimally, potentially leading to training stagnation or slow optimization. To further enhance regression performance, this paper introduces the concept of Inner-IoU and integrates it with MPDIoU, proposing the Inner MPDIoU loss function. By incorporating internal IoU into MPDIoU, the new loss function applies multi-dimensional constraints using corner distances, overlap regions, and internal structure information, resulting in more stable and accurate performance when handling small objects, elongated targets, or objects with complex deformations.

4 Dataset and Evaluations Metrics

4.1 Dataset and Experimental Setup

The dataset used in this study was provided by the State Grid Corporation of China, named the China Power Line Insulator Dataset (CPLID) [32]. This dataset contains 600 images of normal insulators and 248 images of defective insulators, with each image having a resolution of 1152 × 864 pixels. To increase the dataset size and diversity, thereby enhancing the model's generalization ability, multiple data augmentation techniques were applied to the original images. These techniques include geometric transformations, noise addition, contrast adjustment, brightness variation, and saturation modification. Geometric transformations were employed to increase sample diversity from different shooting angles, improving the model's ability to detect defects in tilted or skewed insulators. Gaussian noise was added to simulate potential image disturbances during UAV image capture, thereby enhancing the model's ability to recognize small defects under noisy conditions. Brightness adjustments were applied to simulate different lighting conditions, such as strong sunlight or shadows, enhancing the model's perception of defect details under varying illumination. Using these data augmentation techniques, the dataset was expanded to a total of 3409 images. Subsequently, the dataset was split into training, validation, and test sets in an 8:1:1 ratio. Fig. 8 shows examples of insulator images from the CPLID dataset.



Figure 8: The sample images from the CPLID dataset.

The experiments were conducted on a Linux operating system. The computing platform was equipped with an NVIDIA A40 GPU with 48 GB of memory. The software environment included Python 3.8.12 as the programming language and PyTorch 1.12.1 as the deep learning framework, combined with CUDA 11.3 and the cuDNN acceleration library to improve training and inference efficiency. During training, the input image size was set to 640 × 640, the batch size was 16, the number of data loading threads was 16, and mixed-precision training was enabled to reduce memory usage and accelerate training. The optimizer used was stochastic gradient descent (SGD) with an initial learning rate of 0.01. A warmup strategy was applied during the first 3 epochs to gradually increase the learning rate to its initial value, followed by a cosine decay schedule that reduced the learning rate to 0.01 × 0.01 of the initial value to ensure training stability. The total number of training epochs was 300, with an early stopping strategy set to patience = 100 to prevent overfitting.

4.2 Evaluation Metrics

This paper uses five evaluation metrics—Precision, Recall, mean Average Precision (mAP), number of Parameters (Params), and computational complexity (GFLOPs)—to assess the performance of the proposed model in the task of defect detection on transmission line insulators. Among them, Precision measures the proportion of predicted positive samples that are truly positive, reflecting the model's accuracy; Recall measures the proportion of true targets that are correctly detected, indicating the model's coverage ability; mAP, as a comprehensive metric, more fully reflects the overall detection performance of the model across different categories and IoU thresholds, and is a key standard for evaluating detection performance. In addition, this paper introduces the number of Parameters (Params) and computational complexity (GFLOPs). The former indicates the structural complexity of the model and indirectly reflects its storage and memory requirements; the latter measures the computational cost during inference, serving as an important basis for evaluating the model's real-time performance. The definitions of Precision, Recall, and mean Average Precision are as follows:

$$P = \frac{TP}{TP + FP} \tag{9}$$

$$R = \frac{TP}{TP + FN} \tag{10}$$

$$AP = \int_0^1 p(r)\, dr \tag{11}$$

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \tag{12}$$

where TP represents the number of samples predicted as positive and actually positive, FP represents the number of samples predicted as positive but actually negative, and FN represents the number of samples predicted as negative but actually positive. $p(r)$ denotes the precision curve at different recall levels r, N is the number of target classes, and is the average precision for the i class.

5 Experiments and Results

5.1 Ablation Study

To verify the actual contribution of each module to the model performance, a series of ablation experiments were designed to evaluate the impact of various improvement strategies on the defect detection task of transmission line insulators from the perspective of different module combinations. The detailed experimental settings and results are shown in Table 1.

Table 1: Ablation study on AMST-YOLO.

ADown	Triplet Attn	Inner MPDIOU	CMSC	Precision (%)	Recall (%)	mAP@0.5 (%)	Params (M)	GFLOPs (G)
–	–	–	–	90.0	89.2	91.1	2.5	6.3
✓	–	–	–	90.9	90.6	92.8	2.0	5.1
–	✓	–	–	93.0	90.8	93.1	2.5	6.0
✓	✓	–	–	90.6	89.6	92.6	2.0	4.8

(Continued)

Table 1 (continued)

ADown	Triplet Attn	Inner MPDIOU	CMSC	Precision (%)	Recall (%)	mAP@0.5 (%)	Params (M)	GFLOPs (G)
✓	–	✓	–	96.2	93.2	94.7	2.0	5.1
✓	✓	–	✓	91.2	89.9	92.5	2.0	5.0
✓	✓	✓	–	95.7	93.7	95.2	2.0	4.8
✓	✓	✓	✓	97.0	94.7	96.3	2.0	5.0

In the defect detection of insulators, small defects are prone to being missed due to their limited pixel representation, while closely spaced or partially occluded defects can result in inaccurate localization. Furthermore, complex backgrounds often introduce noise, increasing the likelihood of false detections and reducing the model's precision. To address redundancy in feature extraction and improve computational efficiency, the ADown module was introduced. By performing downsampling while preserving critical feature information, this module reduces the number of parameters and computational cost by approximately 22% and 20%, respectively, while increasing mAP by 1.7%, demonstrating that it can reduce redundant computations without compromising feature representation. Even with reduced redundancy, accurately identifying key target regions—especially small or occluded defects—remains challenging in complex scenes. To enhance the model's focus on these critical areas, the Triplet Attention module was incorporated. This module adaptively weights spatial and channel features, emphasizing important regions while suppressing background interference. When applied alone, it improved precision by 2% and recall by 1.6%, indicating that the model's ability to detect small or occluded defects in complex backgrounds was effectively enhanced. Moreover, traditional IoU-based loss functions have limitations when handling irregularly shaped or densely arranged defects, often resulting in imprecise bounding box regression. To address this, the Inner MPDIoU loss function was introduced, combining boundary point distances and internal region constraints to generate boxes that more closely match the true shapes and positions of defects. When used in combination with ADown, precision increased by 6.2%, recall by 4%, and mAP@0.5 reached 94.7%, demonstrating that more accurate localization directly improves detection performance. Finally, integrating all modules into the YOLO11 framework produced synergistic benefits. Precision increased by 7%, recall by 5.5%, and mAP@0.5 by 5.2%, while the number of parameters and computational cost decreased by 22% and 20%, respectively. This demonstrates that the combined design effectively addresses the main challenges in insulator defect detection—reducing redundant computations, enhancing attention to key regions, and improving bounding box regression—leading to a significant improvement in overall detection accuracy, robustness, and efficiency, especially for small, occluded, or densely arranged defects in complex environments.

5.2 Comparative Experiments and Analysis

5.2.1 Comparison of Different Algorithms

To comprehensively evaluate the overall performance of the proposed model in the task of insulator defect detection, this study conducts comparative experiments under identical experimental conditions using a variety of representative object detection models. Specifically, several models designed for insulator defect detection are included, such as Lite-YOLO-ID [16], FINet [17], CACS-YOLO [18], and YOLO-ELA [19]. In addition, to further verify the advantages of the proposed model in handling complex background interference and preserving small-target features, two improved detection models related to

the challenges addressed in this work—YOLOn-CBAM [33] and Faster-YOLO11 [34]—are incorporated. The former enhances feature discrimination under complex backgrounds through an attention mechanism, while the latter improves small-object detection by strengthening multi-scale feature fusion. Moreover, widely adopted general-purpose models such as YOLOv12 [35], YOLOv13 [36], and YOLO11, as well as more complex yet higher-accuracy models including Gold YOLO [37] and RT-DETR [38], are selected as additional performance baselines. The experimental results are presented in Table 2.

Table 2: Comparative results of mainstream object detection models and the proposed method on the dataset.

Model	Precision/%	Recall/%	mAP@0.5/%	mAP@0.5:0.95/%	Params/M	GFLOPs/G	Model Size/MB
YOLOv12	90.0	88.9	91.4	65.0	2.50	5.8	5.5
YOLOv13	90.9	89.8	92.6	65.5	2.44	6.2	5.4
YOLO11n-CBAM	91.9	91.0	91.9	63.5	2.73	6.3	5.8
YOLO11	90.0	89.2	91.1	64.3	2.58	6.3	5.2
Faster-YOLO11	93.3	92.2	93.2	62.9	5.95	16.0	12.2
Gold-YOLO	92.4	74.4	92.4	63.8	5.61	12.07	11.9
RT-DETR	92.1	93.1	95.6	61.9	31.9	103.4	66.2
Lite-YOLO-ID	95.6	94.3	95.7	63.0	3.73	9.4	7.9
FINet	76.4	87.6	86.8	55.4	21.61	51.6	43.6
CACS-YOLO	91.6	89.6	91.4	65.0	2.64	7.0	5.5
YOLO-ELA	92.3	91.8	94.1	64.6	3.01	8.1	6.3
Our	97.0	94.7	96.3	61.5	2.0	5.0	4.3

The experimental results demonstrate that the proposed AMST-YOLO model achieves the best overall performance in the task of insulator defect detection. Specifically, the model reaches 97.0% in Precision and 94.7% in Recall, outperforming all compared methods. This indicates that AMST-YOLO can accurately identify insulator defects while effectively reducing false positives and false negatives. In terms of the core metric mAP@0.5, the proposed method achieves 96.3%, which is 4.9%, 3.7%, and 5.2% higher than the mainstream YOLO models YOLOv12, YOLOv13, and YOLO11, respectively. These improvements confirm that the model's design effectively addresses challenges such as small target detection, dense or partially occluded defects, and complex background interference.

Compared with insulator-specific improved models such as Lite-YOLO-ID, CACS-YOLO, and YOLO-ELA, although AMST-YOLO does not achieve the highest mAP@0.5:0.95, it consistently outperforms in Precision, Recall, and mAP@0.5. For example, Lite-YOLO-ID achieves an mAP@0.5 close to AMST-YOLO but has 1.73 M more parameters and 4.4 GFLOPs higher computational cost, which reduces its suitability for lightweight deployment on UAVs or embedded devices. Similarly, CACS-YOLO improves feature representation through channel attention, but its Precision (91.6%) and Recall (89.6%) remain lower than AMST-YOLO, indicating less effective suppression of background interference and weaker small-target perception. Specifically, lower Precision means the model tends to produce more false positives, often misclassifying background clutter as small insulator defects, while lower Recall indicates that actual small defects are missed. Together, these metrics quantitatively reflect that the model struggles to accurately detect small targets in complex or noisy environments.

For high-performance yet computationally heavy models such as RT-DETR and Gold YOLO, although RT-DETR reaches a comparable mAP@0.5 (95.6%), its mAP@0.5:0.95 drops to 61.9%, reflecting instability across IoU thresholds due to less accurate localization of small or overlapping targets. Moreover, its parameter count (31.9 M) and computational cost (103.4 GFLOPs) severely limit practical deployment. Gold YOLO, while achieving relatively high Precision (92.4%), has a low Recall (74.4%), showing significant missed detections. In contrast, AMST-YOLO achieves both high Precision and high Recall, demonstrating that the

combination of ADown, Triplet Attention, CMSC, and Inner MPDIoU loss enables more accurate feature extraction, focused attention on key target regions, and improved bounding box regression. Quantitatively, AMST-YOLO reduces redundant computations by 22% and computational cost by 20% compared with the baseline YOLO11, while improving mAP@0.5 by 5.2%. These results confirm that the proposed model not only enhances detection accuracy but also maintains a lightweight design suitable for real-world deployment, effectively addressing the challenges of small target perception, complex backgrounds, and dense or occluded defects.

5.2.2 Comparison Experiments of Partial Improvement Modules

To verify the effectiveness of the Triplet Attention module, it was compared with several popular attention mechanisms, including the Simple Attention Module (SimAM) [39], Global Attention Mechanism (GAM) [40], Coordinate Attention (CA) [41], Large Separable Kernel Attention (LSKA) [42], and the commonly used CBAM. The results are shown in Table 3.

Table 3: Comparison of experimental results of different attention mechanisms.

Algorithm and Module	Precision/%	Recall/%	mAP@0.5/%	Param/M	GFLOPs/G	mAP@0.5:0.95/%
YOLO11	90.0	89.2	91.1	2.58	6.3	64.3
YOLO11-SimAM	89.5	88.2	90.0	2.50	5.9	63.8
YOLO11-LSKA	90.6	89.2	91.9	2.75	6.1	65.9
YOLO11-GAM	90.1	89.4	91.5	4.14	7.2	64.1
YOLO11-CA	91.0	90.3	91.6	2.51	5.9	64.5
YOLO11-CBAM	88.7	88.9	91.0	2.57	6.0	63.9
YOLO11-Triplet Attention	93.0	90.8	93.1	2.5	6.0	64.9

Insulator defect detection is challenged by complex backgrounds, such as surrounding structures or environmental noise, which often lead to false positives and missed detections. To address this issue, the Triplet Attention module was introduced, which jointly models spatial and channel interactions to adaptively highlight regions likely to contain defects while suppressing irrelevant features. When integrated into YOLO11, Triplet Attention improves mAP@0.5 to 93.1%, a 2.0% increase over the baseline, without significantly increasing model complexity (2.5 M parameters and 6.0 GFLOPs).

Compared with other attention mechanisms, Triplet Attention performs better due to the following reasons: LSKA uses large-kernel convolutions to capture both local and global context, but its higher parameter count increases computational burden, making it less practical for lightweight deployment. GAM effectively combines channel and spatial information but enlarges the model size (4.14 M parameters), which may hinder real-time applications. CA is lightweight but does not sufficiently emphasize defect regions under complex backgrounds. SimAM and CBAM achieve moderate results and are less effective at distinguishing defect-relevant features from background noise. In contrast, Triplet Attention explicitly strengthens the model's focus on potential defect regions while suppressing irrelevant background.

Thus, the superiority of Triplet Attention arises from its ability to balance accuracy and efficiency: it improves recognition of defect-relevant regions without increasing computational cost, reduces false positives in complex scenes, and maintains high recall. This explains why it consistently outperforms other attention mechanisms in practical insulator defect detection scenarios.

5.2.3 Comparison Experiment on Other Dataset

To further evaluate the generalization ability of the proposed method, we conducted experiments on the publicly available insulator-defect dataset [43] and compared our method with other insulator defect detection models. The experimental results are shown in the Table 4.

Table 4: Comparison of experimental results of different detection models on the insulator-defect dataset.

Model	Precision/%	Recall/%	mAP@0.5/%	Param/M	GFLOPs/G	mAP@0.5:0.95/%	Model Size/MB
YOLO11	90.2	83.6	87.9	2.58	6.3	55.7	5.5
YOLOv13	91.0	80.0	87.1	2.44	6.2	51.9	5.5
YOLOv12	87.6	81.8	85.3	2.50	5.8	55.1	5.5
Lite-YOLO-ID	90.2	82.2	86.6	3.73	9.4	53.6	7.9
FINet	70.0	91.4	91.3	21.6	51.6	56.3	43.6
CACS-YOLO	90.9	81.6	86.9	2.65	7.1	55.8	5.6
Our	90.7	85.7	89.4	2.08	5.0	55.8	4.2

The insulator-defect dataset contains a total of 1578 images, including three classes: normal insulators, damaged insulators, and flashover insulators. The dataset is divided into training, validation, and test sets according to a 7:2:1 ratio. Considering storage limitations, we adopted online data augmentation during training, where each image is randomly augmented when loaded. The augmentation strategies include horizontal flipping, random translation, random scaling, and random adjustments in the HSV color space, including saturation and brightness modifications. These augmentations increase the diversity of the training samples, improving the model's generalization ability and robustness on small datasets.

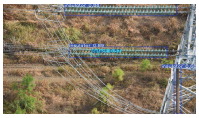
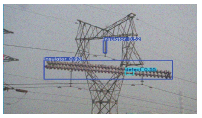
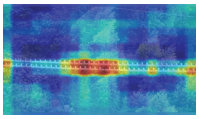
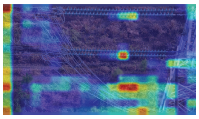
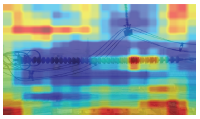
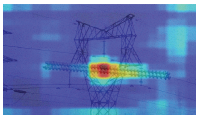
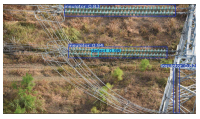
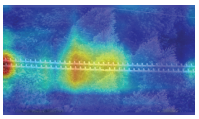
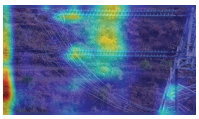
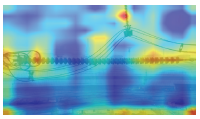
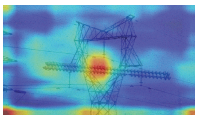
The experimental results show that our method achieves Precision 90.7%, Recall 85.7%, mAP@0.5 89.4%, and mAP0.5:0.95 55.8% on the insulator-defect dataset. Although the Precision of our method is slightly lower than YOLOv13 (91%), its Recall and mAP@0.5 are higher, indicating more stable performance in detecting a larger number of targets and overall recognition capability. Meanwhile, although our method's Recall and mAP metrics are slightly lower than FINet, FINet's model parameters (21.6 M), computational cost (51.6 GFLOPs), and model size (43.6 MB) are much larger than ours. In contrast, our method has the advantage of a lightweight model, making it more suitable for practical deployment and fast detection tasks.

5.3 Visualization of Experimental Results

To more intuitively evaluate the adaptability and stability of the proposed AMST-YOLO model under different detection environments, Table 5 presents a comparison of detection results and heatmap visualizations before and after the improvements across various scenarios. The selected scenarios cover a wide range of challenges commonly encountered in transmission line insulator defect detection tasks, including large targets in complex backgrounds (Scenario 1), small targets in complex backgrounds (Scenario 2), small targets in simple backgrounds (Scenario 3) and scenes with introduced noise interference (Scenario 4). The heatmap visualizations provide a direct reflection of the model's attention distribution across these scenarios, where deeper colors indicate higher focus by the model on those regions. More concentrated dark areas suggest stronger feature extraction in those regions. The comparison reveals that YOLO11 exhibits relatively scattered attention areas in complex backgrounds or when detecting small targets, resulting in less focused feature extraction and unstable performance, especially for small targets. In contrast, AMST-YOLO, through the incorporation of attention mechanisms and an improved bounding box optimization

strategy, enhances the model's ability to concentrate on critical regions, effectively suppresses background interference, and improves detection accuracy and robustness. The heatmap results further confirm the effectiveness of these improvements: the heatmaps generated by AMST-YOLO are more focused on the target regions, demonstrating superior feature extraction capability and attention allocation strategy, thereby exhibiting stronger detection stability and generalization ability in complex environments.

Table 5: Comparison of target detection results and heatmap visualizations before and after improvements in different scenarios.

Item	Scenarios			
	Scene 1	Scene 2	Scene 3	Scene 4
Original Image				
YOLO11				
YOLO11 Heatmap				
Ours				
Our Heatmap				

6 Conclusion

To address the challenges of weak small-object perception, complex backgrounds, and high model complexity in transmission line insulator defect detection, this paper proposes the AMST-YOLO detection algorithm. The approach involves designing a CMSC module to optimize feature representation, introducing a lightweight ADown sampling module and triplet attention mechanism to enhance feature extraction, and employing an improved MPDIoU loss function based on the Inner concept to boost localization accuracy. Experimental results demonstrate that the proposed algorithm achieves a 5.2% improvement in mean average precision while reducing computational cost by 20.6% to 5.0 GFLOPs and decreasing the number of parameters by approximately 22%, maintaining model lightweight while significantly enhancing detection performance.

The model still presents room for further optimization, particularly in terms of the stricter evaluation metric mAP@0.5:0.95, which remains relatively insufficient at 61.5%. The possible causes of this scenario include slight deviations in predicted bounding boxes at fine-grained scales, reduced IoU due to closely spaced or partially occluded insulator targets, and the inherent regression limitations introduced by lightweight network design. To address these issues, future work will focus on optimizing the detection head structure, incorporating higher-resolution feature fusion strategies, adopting multi-stage regression mechanisms, and exploring more IoU-sensitive loss functions to further improve localization performance

under strict evaluation conditions. Moreover, AMST-YOLO outperforms other comparison models in terms of model size (4.2 MB), parameter count, and computational cost, achieving significant lightweight design. This makes it suitable for deployment on storage-limited embedded devices or UAV platforms, while also reducing transmission and update overhead. However, lightweight design does not necessarily translate to faster inference. The introduction of the CMSC module, ADown downsampling, and Triplet Attention mechanism to enhance feature extraction adds extra computational overhead during inference, resulting in an average FPS of 111.6, slightly lower than the original YOLO11 at 125.9. Future work will focus on further optimizing these modules to improve inference speed while maintaining high accuracy and the advantages of a lightweight model.

Acknowledgement: We thank the Master's student Haonan Zhang and Zhengxin Zhan for their helpful research suggestions.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization: Yanping Chen, Le Zou, Hailei Wang; data collection: Ran Guo, Junchao Liu, Zailu Shen; analysis and interpretation of results: Yanping Chen, Le Zou, Hailei Wang; draft manuscript: Ran Guo. All authors reviewed and approved the final version of the manuscript.

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

Ethics Approval: Not applicable.

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

References

1. Liu Y, Liu D, Huang X, Li C. Insulator defect detection with deep learning: a survey. IET Generat Transm Distrib. 2023;17(16):3541–58. doi:10.1049/gtd2.12916.
2. 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 Instrument Measur. 2022;71:1–12. doi:10.1109/TIM.2022.3200861.
3. Feng Z, Guo L, Huang D, Li R. Electrical insulator defects detection method based on yolov5. In: 2021 IEEE 10th Data Driven Control and Learning Systems Conference (DDCLS). Piscataway, NJ, USA: IEEE; 2021. p. 979–84. doi:10.1109/DDCLS52934.2021.9455519.
4. Li Y, Ni M, Lu Y. Insulator defect detection for power grid based on light correction enhancement and YOLOv5 model. Energy Rep. 2022;8(1):807–14. doi:10.1016/j.egyr.2022.08.027.
5. Liu J, Hu M, Dong J, Lu X. Summary of insulator defect detection based on deep learning. Elect Power Syst Res. 2023;224:109688. doi:10.1016/j.epsr.2023.109688.
6. Bai X, Xie Y, Zhao M, Wu H, Zhang W, Tan Y, et al. Few-shot defect detection of insulators based on local feature depth information. Power Syst Technol. 2024;48(2):740–9. doi:10.13335/j.1000-3673.pst.2022.2187.
7. X. LI, H. SU, G. LIU. Insulator defect recognition based on global detection and local segmentation. IEEE Access. 2020;8:59934–46. doi:10.1109/ACCESS.2020.2982288.
8. Waleed D, Mukhopadhyay S, Tariq U, El-Hag AH. Drone-based ceramic insulators condition monitoring. IEEE Trans Instrum Meas. 2021;70:1–12. doi:10.1109/TIM.2021.3078538.
9. Tang J, Wang J, Wang H, Wei J, Wei Y, Qin M. Insulator defect detection based on improved faster R-CNN. In: 2022 4th Asia Energy and Electrical Engineering Symposium (AEEES). Piscataway, NJ, USA: IEEE; 2022. p. 541–6. doi:10.1109/AEEES54426.2022.9759683.

10. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In: IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE; 2016. p. 779–88. doi:10.1109/CVPR.2016.91.
11. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C, et al. SSD: single shot multibox detector. In: Computer Vision–ECCV 2016:14th European Conference; 2016 Oct 11–14; Amsterdam, The Netherlands. Cham, Switzerland: Springer. p. 21–37.
12. Lin TY, Goyal P, Girshick R, He M, Dollar P. Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision. Piscataway, NJ, USA: IEEE; 2017. p. 2980–8. doi:10.1109/TPAMI.2018.2858826.
13. Wang Z, Liu X, Peng H, Zheng L, Gao J, Bao Y. Railway insulator detection based on adaptive cascaded convolutional neural network. IEEE Access. 2021;9:115676–86. doi:10.1109/ACCESS.2021.3105419.
14. Wu J, Deng Q, Xian R, Tao X, Zhou Z. An instance segmentation method for insulator defects based on an attention mechanism and feature fusion network. Appl Sci. 2024;14(9):3623. doi:10.3390/app14093623.
15. Jiang T, Wang Y, Dou W. Lightweight ACCP-YOLOv8 based multi-defect detection of insulators. Communicat Inform Technol. 2025;1:16–22. doi:10.1109/ACCESS.2025.3539831.
16. Li D, Lu Y, Gao Q, Li X, Yu X, Song Y. LiteYOLO-ID: a lightweight object detection network for insulator defect detection. IEEE Trans Instrument Measur. 2024;73:1–12. doi:10.1109/TIM.2024.3418082.
17. Zhang ZD, Zhang B, Lan ZC, Liu H, Li D, Pei L, et al. FINet: an insulator dataset and detection benchmark based on synthetic fog and improved YOLOv5. IEEE Trans Instrum Meas. 2022;71:6006508. doi:10.1109/TIM.2022.3194909.
18. Cao Z, Chen K, Chen J, Chen Z, Zhang M. CACS-YOLO: a lightweight model for insulator defect detection based on improved YOLOv8m. IEEE Trans Instrument Measur. 2024;73:1–10. doi:10.1109/TIM.2024.3453332.
19. Akindele O, Atolagbe J. YOLO-ELA: efficient local attention modeling for high-performance real-time insulator defect detection. arXiv:2410.11727.2024.
20. Zhang X, Zhang Y, Hu M, Ju X. Insulator defect detection based on YOLO and SPP-Net. In: 2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE). Piscataway, NJ, USA: IEEE; 2020. p. 403–7. doi:10.1109/ICBASE51474.2020.00092.
21. Ji Y, Zhang D, He Y, Zhao J, Duan X, Zhang T. Improved YOLO11 algorithm for insulator defect detection in power distribution lines. Electronics. 2025;14(6):1201. doi:10.3390/electronics14061201.
22. Shen P, Mei K, Cao H, Zhao Y, Zhang G. LDDFSF-YOLO11: a lightweight insulator defect detection method focusing on small-sized features. IEEE Access. 2025;13:90273–92. doi:10.1109/ACCESS.2025.3569970.
23. Lan X, Wang X, Liu L. RMC-YOLO: an object detection model for insulator defects. Signal ImageVideo Process. 2025;19(12):1–13. doi:10.1007/s11760-025-04568-9.
24. Liang Y, Yang L, Sun S, Li Z, Shi Y, Zhang Z, et al. YOLOv11-RAH: a recurrent attention-enhanced edge intelligence network for UAV-based power transmission line insulator inspection. Int J Intell Netw. 2025;6(2):244. doi:10.1016/j.ijin.2025.11.008.
25. Wang CY, Yeh IH, Mark Liao HY. Yolov9: learning what you want to learn using programmable gradient information. In: European Conference on Computer Vision, 2024. Cham, Switzerland: Springer; 2024. p. 1–21.
26. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE; 2018. p. 7132–41. doi:10.1109/TPAMI.2019.2913372.
27. Woo S, Park J, Lee JY, Kweon I. Cbam: convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV 2018). Cham, Switzerland: Springer; 2018. p. 3–19.
28. Misra D, Nalamada T, Arasanipalai AU, Hou Q. Rotate to attend: convolutional triplet attention module. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Piscataway, NJ, USA: IEEE; 2021. p. 3138–47. doi:10.1109/WACV48630.2021.00318.
29. Li J, Wen Y, He L. Scconv: spatial and channel reconstruction convolution for feature redundancy. In: Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE; 2023. p. 6153–62. doi:10.1109/CVPR52729.2023.00596.
30. Guo MH, Lu CZ, Hou Q, Liu Z, Cheng M, Segnext Hu S. Rethinking convolutional attention design for semantic segmentation. Adv Neural Inform Process Syst. 2022;35:1140–56.

31. Ma S, Xu Y. Mpdiou: a loss for efficient and accurate bounding box regression. arXiv:2307.07662. 2023.
32. 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 Systems Man Cybernet Syst. 2020;50(4):1486–98. doi:10.1109/TSMC.2018.2871750.
33. Niu Z, Zhao Y. A tea buds detection algorithm based on improved YOLO11. In: 2024 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML); 2024 Nov 22–24; Shenzhen, China. p. 1550–3. doi:10.1109/ICICML63543.2024.10958012.
34. Ma X. Dangerous driving behavior detection algorithm based on Faster-YOLO11. Inform Technol Informatiz. 2025;3:175–8. doi:10.3969/j.issn.1672-9528.2025.03.042.
35. Tian Y, Ye Q, Doermann D. Yolov12: attention-centric real-time object detectors. arXiv:2502.12524. 2025.
36. Lei M, Li S, Wu Y, Hu H, Zhou Y, Zheng X, et al. YOLOv13: real-time object detection with hypergraph-enhanced adaptive visual perception. arXiv:2506.17733. 2025.
37. Wang C, He W, Nie Y, Guo J, Liu C, Han K, et al. Gold-YOLO: efficient object detector via gather-and-distribute mechanism. Adv Neural Inform Process Syst. 2023;36:51094–5112.
38. Zhao Y, Lv W, Xu S, Wei J, Wang G, Dang Q, et al. Detrs beat yolos on real-time object detection. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE; 2024. p. 16965–74. doi:10.1109/CVPR52733.2024.01605.
39. Yang L, Zhang RY, Li L, Xie X. Simam: a simple, parameter-free attention module for convolutional neural networks. In: Proceedings of the 38th International Conference on Machine Learning. London, UK: PMLR; 2021. p. 11863–74.
40. Liu Y, Shao Z, Hoffmann N. Global attention mechanism: retain information to enhance channel-spatial interactions. arXiv:2112.05561. 2021.
41. Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE; 2021. p. 13713–22. doi:10.1109/CVPR46437.2021.01350.
42. Lau KW, Po L-M, Rehman YAU. Large separable kernel attention: rethinking the large kernel attention design in cnn. Expert Syst Appl. 2024;236:121352. doi:10.1016/j.eswa.2023.121352.
43. Trash Detection. insulator-defect Dataset [Internet]. Roboflow Universe; 2025 [cited 2025 Nov 21]. Available from: https://universe.roboflow.com/trash-detection-zqfed/insulator-defect-gcnfq.