ITC 3/55
Information Technology and Control
Vol. 55 / No. 3/ 2026
pp. 1190-1204
DOI 10.5755/j01.itc.55.3.42278

FVS-YOLO: A Lightweight Insulator Material Detection Network with Enhanced Feature Representation for UAV-based Power Line Inspection

Received 2025/07/18 | Accepted after revision 2026/06/27

HOW TO CITE: Yao, X., Chai, Y., Li, L., Sun, Z., Liu, W., Liu, Q. (2026) FVS-YOLO: A Lightweight Insulator Material Detection Network with Enhanced Feature Representation for UAV-based Power Line Inspection. *Information Technology and Control*, 55(3), 1190-1204. https://doi.org/10.5755/j01.itc.55.3.42278

FVS-YOLO: A Lightweight Insulator Material Detection Network with Enhanced Feature Representation for UAV-based Power Line Inspection

Xiaomin Yao*, Yujiao Chai, Long Li, Zenglin Sun, Wenshuai Liu, Qiang Liu
School of Mechanical Engineering, Shaanxi University of Technology, Hanzhong, China

Corresponding author: yaoxiaomin@snut.edu.cn

To address the high computational complexity of existing deep learning models in transmission line insulator material detection and their difficulty in deployment on mobile platforms, this paper proposes FVS-YOLO, a lightweight model. Building upon the YOLOv8s architecture, this model implements three key optimizations: First, an improved FasterNet module is introduced into the backbone network, where embedded convolutional layers with stride=1 enhance feature extraction capability while significantly reducing floating point operations (FLOPs). Second, a novel VoVGSCSPC neck module is designed, integrating group convolution and spatial convolution strategies to minimize computational redundancy during feature fusion. Third, the parameter-free SimAM attention mechanism is incorporated to amplify critical features and suppress irrelevant information. Experiments were conducted on a self-built dataset comprising 3,089 images across three insulator material types (composite, glass, and ceramic). Results demonstrate that FVS-YOLO achieves a mAP@50 of 0.916 while reducing parameters to 5.0M (54.95% decrease), computational cost to 15.1 GFLOPs (46.83% reduction), and model size to 9.8MB (54.21% compression). Moreover, its inference speed reaches 229.9 FPS, representing an 18% improvement. This model achieves a favorable balance between accuracy and lightweight design for insulator material detection, showing potential for deployment on drone-edge computing devices and providing a promising solution for real-time smart grid inspection.

KEYWORDS: YOLO; Lightweight; Insulator; FasterNet network structure; SimAM

1. Introduction

Insulators serve as critical components in power systems, functioning to support and isolate live conductors while preventing current leakage. These devices are extensively employed in transmission lines, substations, and various electrical installations. However, insulators are vulnerable to multiple degradation mechanisms including surface contamination, moisture absorption, cracking, material aging, and corrosive deformation due to prolonged exposure to harsh environmental conditions [9]. The detection of such defects is particularly challenging given the typically small dimensions of insulator surface flaws. As fundamental elements of transmission infrastructure, insulators are constantly subjected to environmental stresses that may compromise their integrity [9]. The material composition of insulators constitutes a crucial factor in transmission line performance and reliability. Consequently, the development of efficient and accurate methods for insulator material identification has emerged as a significant research challenge in power system maintenance [1]. The choice of insulator material directly determines several key performance characteristics, including dielectric properties, mechanical robustness, and environmental durability. Contemporary power systems primarily utilize three classes of insulator materials: ceramic, glass, and polymer composites. Each material category exhibits distinct advantages and limitations in terms of electrical performance, mechanical strength, and service life. The appropriate selection and maintenance of insulator materials are therefore essential for ensuring the safe and efficient operation of transmission networks.

Traditional insulator inspection primarily relies on manual patrols. However, this approach faces growing challenges in meeting the demands of modern power grid inspections due to numerous inspection points, extensive coverage areas, complex working environments, as well as high time consumption and labor costs [8]. Currently, unmanned aerial vehicles (UAVs) equipped with real-time image processing systems have emerged as the mainstream solution for overhead transmission line inspections. These systems leverage advantages such as compact size, operational flexibility, low cost, and high efficiency [17], [24]. Specifically, UAVs can capture high-definition images of transmission lines in real time and utilize object detection algorithms to accurately identify insulator material types. This capability enables efficient and intelligent inspection of power transmission infrastructure.

With the rapid development of deep learning technology, current research has focused on integrating object detection algorithms with UAV systems to improve inspection efficiency, making fast and efficient insulator material detection highly significant for practical applications. Current object detection algorithms can be divided into two categories: single-stage detectors and two-stage detectors. Single-stage algorithms including SSD [14] and the YOLO series [10], [19], [20] are renowned for their excellent processing speed. Two-stage detection algorithms such as R-CNN [3], Faster R-CNN [6], and Mask R-CNN [2] achieve higher detection accuracy but face significant challenges in detection speed, requiring improvement compared to single-stage algorithms.

With the continuous advancement of single-stage detection algorithms, numerous researchers have investigated and improved various detection approaches to achieve accurate and rapid insulator identification. Unlike defect detection, insulator material identification is a more challenging fine-grained classification task. Different materials (ceramic, glass, composite) exhibit subtle differences in visual features such as texture, gloss, and color, imposing higher demands on the model's feature extraction capability.

In recent years, some researchers have begun to focus on material-related analysis of insulators. For instance, Limin et al. [16] systematically discuss the chemical composition and structural differences of ceramics, glass, and composite insulators, and analyze the influence of material composition on electrical performance and mechanical properties. Yang et al.'s [13] review focuses on the methods for detecting insulator defects based on the characteristics of sound, light and heat, and covers the differences in the response of different materials (ceramics, glass, composites) to the detection signals. However, the current methods mainly rely on traditional image processing and separate detection, making them

difficult to deploy on resource-constrained UAV platforms for real-time inspection. Research on applying lightweight deep learning models to insulator material identification remains relatively scarce.

In the field of insulator testing, current research mainly focuses on defect identification and material identification [5]. The texture features of insulators are extracted using the Gray-Level Co-occurrence Matrix (GLCM), and the optimal feature subset is selected through the feature selection algorithm for insulator detection and positioning, while Guocui et al. [4] address the influence of insulation insulator image noise on texture feature extraction, two feature descriptors, PHOG (Pyramid Gradient Orientation Histogram) and BOW-SURF (Bag-of-Words Accelerated Robust Feature), which are fused to enhance the ability of expressing detailed textures, but their approach has certain limitations when compared with traditional machine learning and deep learning.

Notwithstanding these improvements in detection accuracy and processing speed, existing methods have yet to effectively address the critical task of insulator material identification – particularly for distinguishing between different material types under practical operating conditions [21]. In the operation and maintenance of power transmission lines, insulators of different materials exhibit distinct aging characteristics, require different inspection standards, and necessitate varied maintenance strategies. Automatic material identification represents a critical step towards achieving refined and intelligent maintenance, yet current research shows a noticeable gap in this area. Under complex natural conditions (such as varying lighting, cluttered backgrounds, and weather interference), simple network pruning often leads to a sharp decline in feature extraction capability – particularly for the subtle textures, gloss, and color features essential for material differentiation [7], [22]. While pursuing speed, existing models struggle to maintain the ability to robustly extract these subtle features against complex backgrounds. Therefore, there is an urgent need for a lightweight and efficient detection method capable of meeting the recognition requirements for different insulator materials.

Current research faces three major challenges in achieving this goal: Firstly, while pursuing model lightweighting and speed improvement, how to effectively maintain or even enhance the detection accuracy for insulator materials remains a critical challenge, as simple network pruning often leads to diminished feature extraction capability and accuracy loss [15], [18]. Secondly, insulator images often contain complex background interference (such as skies, transmission towers, and wires) and noise, requiring the model to possess robust multi-scale feature fusion capabilities and the ability to focus on key features to ensure algorithmic robustness [5], [25]. Lastly, for the specific task of insulator material detection [11], how to reduce model complexity and computational resource consumption (e.g., FLOPs and parameter count) while avoiding memory access becoming a performance bottleneck poses a practical challenge for efficient deployment on edge devices.

To address these challenges, this study implements targeted lightweight improvements based on the YOLOv8s model, aiming to reduce network complexity while ensuring accuracy, thereby enabling rapid and precise detection of insulator materials and contributing to the safety and reliability of power transmission line operation and maintenance. The proposed FVS-YOLO model overcomes the existing challenges through the following three core improvements:

1. Introducing a lightweight FasterNet module to replace the backbone network of the YOLOv8 model. This directly addresses the challenge of balancing computational efficiency and accuracy. The module significantly reduces floating-point operations and frequent memory access through its partial convolution strategy. While achieving network lightweighting, its powerful feature extraction capability effectively mitigates the potential accuracy loss caused by backbone network simplification.
2. Adopting the VoVGSCSPC module to replace the original C2f module in the model's neck. This design aims to tackle the challenges of feature fusion and robustness in complex backgrounds. The VoVGSCSPC module enhances multi-scale fusion capabilities and incorporates attention mechanisms, optimizing the integration of features at different scales and strengthening the extraction and utilization of key insulator features. This improves the model's recognition accuracy in complex environments while further reducing the overall model complexity and computational burden.

3 Embedding the parameter-free SimAM attention mechanism. This improvement specifically addresses the challenge of enhancing the salience of key features without increasing computational costs. Based on neuroscience theory, SimAM uses an energy function to enhance important feature channels and suppress redundant information. It improves the network's focus on key regions of insulators and optimizes feature representation without introducing additional parameters, thereby enhancing the model's robustness to complex backgrounds and image noise.

2. Methods

2.1. YOLOv8

YOLOv8 continues the backbone network and neck network design of CSPDarkNet. It replaces the C3 module of YOLOv5 with the C2f module, further enhancing the lightweight effect. It adopts the design concept of dual detection heads and anchor-free boxes, making the training and inference of the model more efficient. As shown in Figure 1, the network structure of YOLOv8 is presented. It mainly consists of three parts: the backbone, the neck, and the head. The backbone part is composed of 5 convolutional modules, 4 C2f modules, and 1 SPPF module, used for feature extraction. Compared to the C3 module, the C2f module has lower parameters and better feature extraction capabilities. The neck is composed of 4 C2f modules, 2 convolutional modules, and 2 upsampling modules to further process and fuse the features from the backbone part. Through downsampling and upsample operations, the detailed features at the lower level and the semantic features at the higher level are fused together to obtain a more comprehensive and rich feature representation. The head part uses decoupled heads to process the feature maps of three different levels and generate output results. The aim is to achieve more efficient object detection, with stronger feature extraction capabilities compared to YOLOv5, and performs particularly well in complex backgrounds and object detection.

Figure 1
Architectural Overview of YOLOv8 Network.

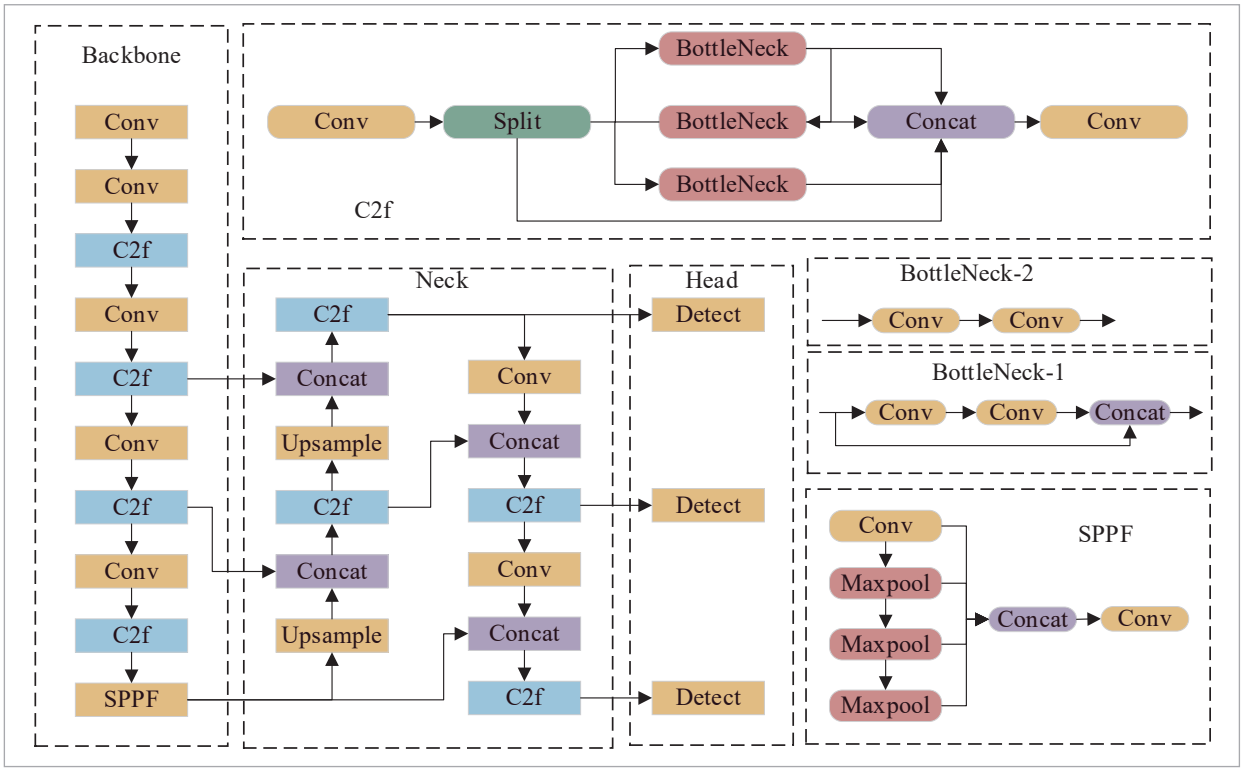
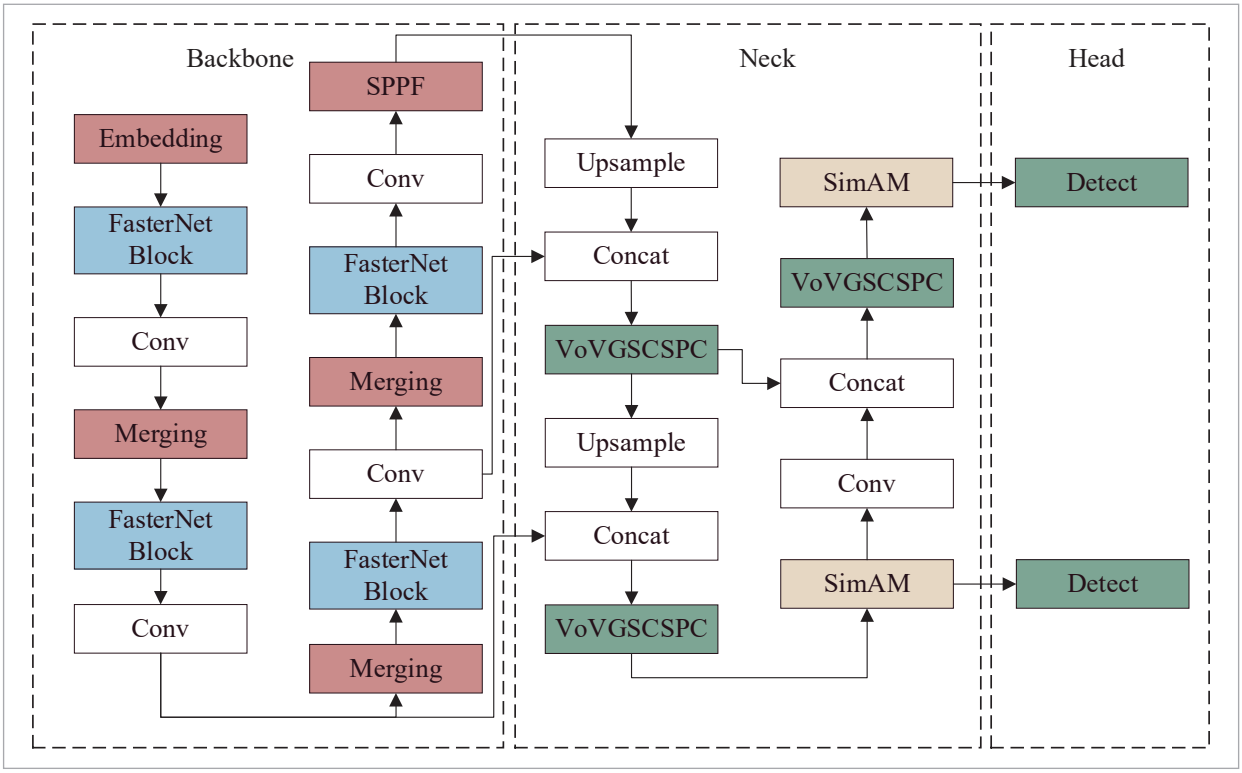


Figure 2
Architectural Diagram of FVS-YOLO Network.



2.2. FVS-YOLO

Although the YOLOv8 model performs well in general image processing, it still poses challenges in detecting insulator materials under complex natural environments and image noise. Therefore, addressing the issues of YOLOv8's slow detection speed, high computational cost, and large model size, this study proposes a lightweight insulator material detection model called FVS-YOLO, aiming to achieve rapid detection while ensuring accuracy, in order to ensure the safe and stable operation of power transmission lines. The model structure diagram is shown in Figure 2. This model introduces an improved FasterNet backbone to replace the original backbone block. This module has stronger feature extraction capabilities, alleviating the reduction in model accuracy caused by the embedding of lightweight backbone networks. Additionally, a VoVGSCSPC module is designed to replace the C2f module of the original model, enabling further reduction in the model's parameters and computational cost. To effectively enhance the performance of neural networks, a SimAM parameter-free attention mechanism module is introduced. This further improves the model's performance.

2.3. Improve the FasterNet Module

In the application of insulation material detection for high-voltage transmission lines, various complex environmental backgrounds, such as changing viewpoints due to fog, result in large computational requirements and oversized models for the original YOLOv8s model. The lightweight design of convolutional neural networks typically relies on strategies such as factorized convolutions and sparse connections, which can significantly reduce model complexity. However, these approaches commonly face the limitation of potentially compromising feature extraction capability, leading to accuracy degradation. To construct a more lightweight and efficient backbone network, this paper proposes an improved FasterNet module as the backbone network. The core innovation of FasterNet lies in its partial convolution (PConv), which reduces redundant computations and

memory access while maintaining strong feature extraction capabilities, thereby achieving an optimal balance between lightweight design and performance.

Material recognition depends on capturing subtle features such as textures and color nuances. Traditional lightweight networks (e.g., MobileNet) employing depthwise separable convolutions may cause feature degradation, whereas FasterNet's PConv strategy better preserves these critical discriminative features. Our selection of FasterNet is precisely aimed at ensuring the extraction of material-discriminative features while maintaining lightweight efficiency.

Building upon the original FasterNet architecture, we embed a convolutional layer with stride=1 after each basic module. This enhancement strengthens the feature extraction capability and compensates for the inherent limitations of lightweight networks in feature representation.The improved FasterNet backbone network is shown in Figure 3. The core innovation lies in the partial convolution (PConv) operation, which achieves lightweight design by reducing redundant computations. Given an input feature map $X \in R^{C\times H\times W}$, where C represents the number of channels, the partial convolution first splits the input along the channel dimension into two partitions as shown in Equation (1):

$$X = [X_A, X_B], \quad X_A \in R^{rC\times H\times W}, X_B \in R^{(C-rC)\times H\times W}, \quad (1)$$

where r represents a partition ratio (typically r=4). The partial convolution performs standard convolution operations only on one partition (e.g., X^A), while leaving the other partition (X^B) unchanged. This process can be formulated as shown in Equation (2):

$$Y_A = \mathrm{Conv}(X_A), \quad Y_B = X_B. \quad (2)$$

Finally, the outputs are concatenated, with an optional channel shuffle operation applied to facilitate information flow between the two partitions, as shown in Equation (3):

$$Y = \mathrm{Shuffle}([Y_A, Y_B]). \quad (3)$$

2.4. VoVGSCSPC Module

Following the optimization of the backbone network, we further implement a lightweight design for the neck network. Specifically targeting the neck architecture of YOLOv8s, which primarily employs C2f modules for feature fusion, we address a critical limitation: while C2f modules effectively integrate features, their dense cross-layer connections and convolutional operations introduce substantial computational redundancy. This inherent characteristic hinders optimal deployment on resource-constrained edge devices. Therefore, this paper uses lightweight network modules to improve the C2f module of the neck network. The improvement is achieved through the GSConv (Group-wise and Spatial Convolution) module, which optimizes the Cross-Stage Partial connection (CSP) structure. By enabling more efficient gradient flow and information fusion, it reduces computational complexity while enhancing feature representation. This module aims to balance the model's representational capacity and computational efficiency through the integration of group convolution and standard spatial convolution.

Insulators exhibit different levels of features at various capture scales. The VoVGSCSPC module is introduced into the neck network to strengthen multi-scale fusion capability, enabling the model to comprehensively leverage both macroscopic shapes and microscopic textures for material identification.

Figure 3
The improved structure of the FasterNet network.

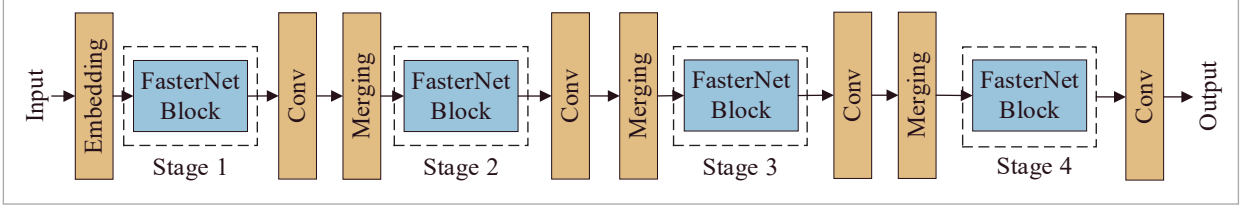
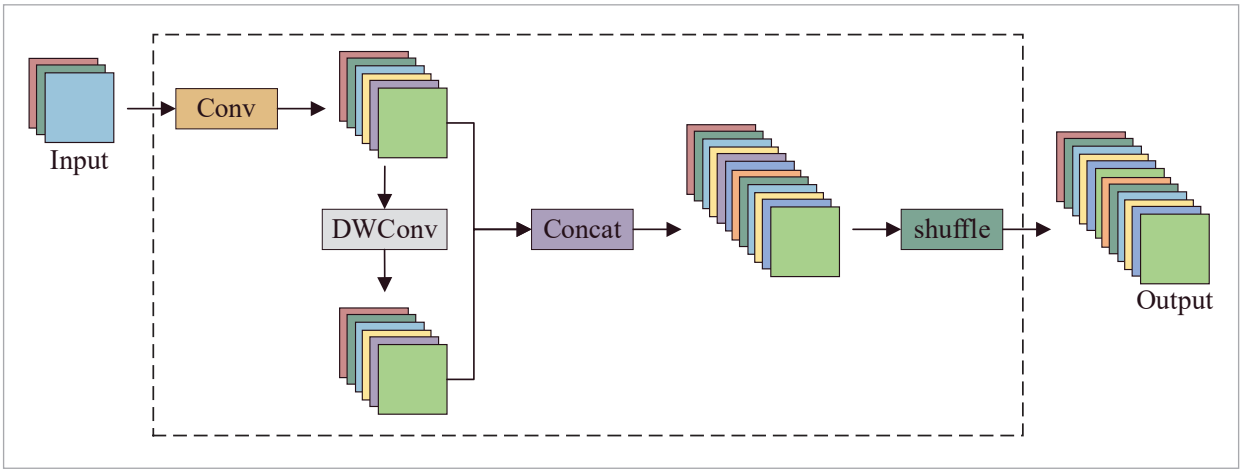


Figure 4
Structure diagram of GSConv network.



Specifically, GSConv first applies group convolution to the input feature map to reduce inter-channel computational complexity. Subsequently, it employs a standard spatial convolution layer to promote information fusion across different feature groups, compensating for potential information loss caused by pure group convolution. This "group-then-fuse" strategy significantly reduces the number of parameters while maintaining rich feature representation capability. The schematic diagram of the GSConv structure is shown in Figure 4. Its core can be regarded as an efficient feature transformation function. Assuming the module has two input branches X_1 and X_2 (e.g., feature maps from different scales), the generation process of its output Y can be simplified as Equation (4):

$$Y = \mathrm{FVoVGSCSPC}(X_1, X_2) = \mathrm{Conv}\left(\mathrm{Cat}[G(X_1), H(X_2)]\right) + X_{\mathrm{shortcut}}. \quad (4)$$

As shown in Figure 5, based on GSConv, we further construct the GS Bottleneck module. Specifically, the GS Bottleneck module consists of two GSConv layers connected via a residual connection for feature reuse. The input features first pass through the first GSConv for channel compression and spatial feature extraction, then through the second GSConv to restore the channel dimension, and finally are added to the input features. This design reduces computational cost while maintaining stable gradient flow.

As illustrated in Figure 6, the VoVGSCSPC module for the cross-stage partial network is designed based on the GS bottleneck module using a one-shot aggregation approach. The VoVGSCSPC module adopts a cross-stage partial network structure. Input features are split into two parts along the channel dimension: one part is directly passed to the output, while the other part sequentially passes through the GS Bottleneck module for feature extraction. The two parts are concatenated along the channel dimension and then fused via GSConv. This "partial processing, partial bypass" design effectively reduces redundant gradient information and computational cost.

Figure 5
Structure of the GS bottleneck module.

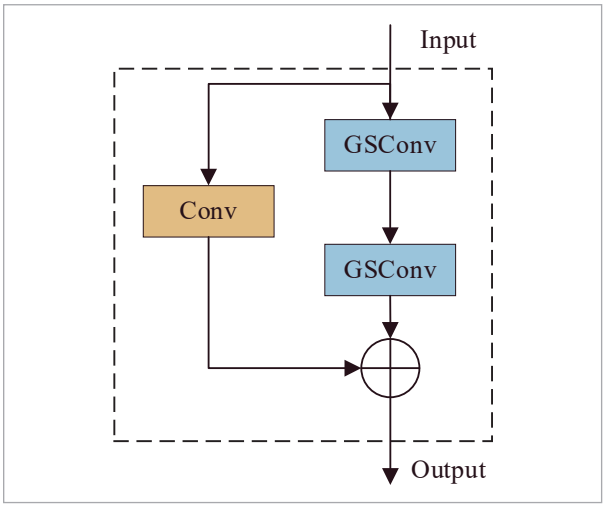
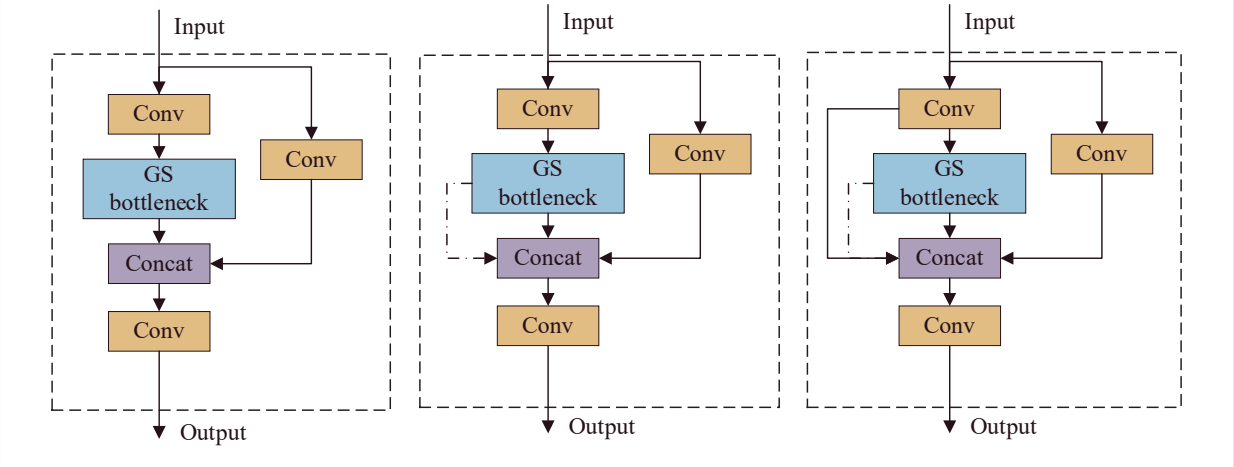


Figure 6
Structure of the VoVGSCSPC module.



2.5. SimAM Unparameterized Attention Module

To enhance the model's ability to focus on key features and suppress complex background interference, this paper embeds the SimAM (Simple Attention Module) parameter-free attention mechanism before the detection head. The core innovation of SimAM lies in its neuroscientifically inspired energy function, which assigns an independent importance weight to each neuron without introducing any learnable parameters.

For an input feature map $X \in R^{C\times H\times W}$, SimAM evaluates the importance of each neuron through the following energy function, as shown in Equation (5):

$$e_t = \frac{4(\hat{\sigma}^2 + \lambda)}{(t-\hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda}, \tag{5}$$

where t is the target neuron, $\hat{\mu}$ and $\hat{\sigma}^2$ are the mean and variance of all neurons except t, and λ is a hyperparameter. A lower energy value e_t indicates higher discriminability between the target neuron and its surrounding neurons, implying greater importance.

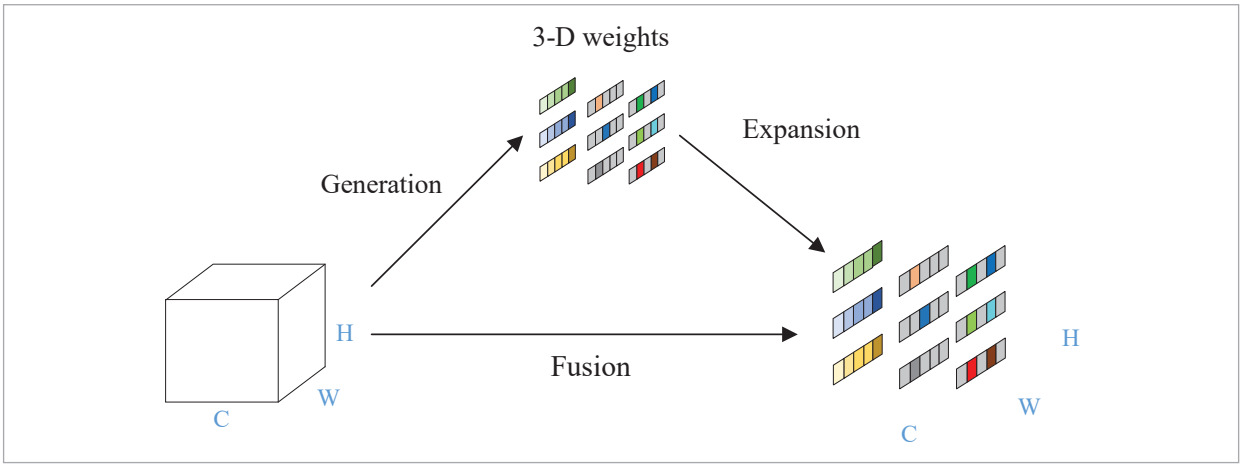
Finally, the weighted feature map produced by SimAM is expressed as shown in Equation (6):

$$\tilde{X} = \text{sigmoid}\left(\frac{1}{E}\right) \odot X, \tag{6}$$

where E is the energy matrix computed by the energy function, and $\odot$ denotes element-wise multiplication. This parameter-free design enables SimAM to adaptively enhance important features and suppress redundant information without increasing model complexity, making it particularly suitable for deployment on resource-constrained edge devices. Figure 7 illustrates the structural diagram of the SimAM attention mechanism.

In addition to structural optimization, to address interference from complex backgrounds, we introduce a mechanism that enables the model to focus on the insulator itself while suppressing background noise, thereby enhancing the model's ability to focus on key features. The SimAM non-parametric attention mechanism is introduced before the detection head to improve focus on critical regions and directly enhance the robustness of material classification. The prominent advantage of SimAM lies in its non-parametric nature and the unified weight calculation based on a neuroscientific energy function. This means it can adaptively enhance important features without increasing model complexity. Its core idea is to assign a unique weight to each neuron in the feature map, derived through a general energy function, thus avoiding the introduction of additional model parameters. This parameter-free design enables SimAM to adaptively enhance key features and suppress redundant information without increasing model complexity, making it particularly suitable for deployment on edge devices with limited computational resources.

Figure 7
Structure Diagram of Parameter-Free Attention Mechanism.



3. Experimental Design and Data Analysis

3.1. Experimental Environment and Configuration

This experiment was conducted on the Windows 10 system. The hardware system for implementing the algorithm was set up with RTX 4060Ti-8GB, 32GB memory, and i7-13700KF. Python 3.10, the deep learning framework Pytorch 2.1.0, and CUDA 11.8 were used. The parameter settings in the experiment are shown in Table 1.

Table 1
Experimental Parameter Settings.

Name	parameters
Input size	640 640
Batch-size	4
Training rounds	250
Initial learning rate	0.01
Optimizer	SGD
Learning rate momentum	0.937
Weight decay coefficient	0.0005
Learning rate scheduler	Cosine Annealing
Confidence threshold	0.25
NMS threshold	0.5
Random seed	0

3.2. Experimental Data Set

Based on the China Transmission Line Insulator Data Set (CPLID), and by obtaining some images from the images captured by unmanned aerial vehicles for the inspection of transmission lines, the insulator material data set and the insulator defect data set were obtained. As shown in Figure 8, the insulator material dataset comprises three categories: composite, glass, and ceramics. The original data was expanded through data augmentation techniques. Data augmentation expands the dataset by applying a series of transformations to the original data, including rotation, brightness adjustment, and noise addition. The initial insulator dataset was expanded from 2,197 images to 3,089 images, effectively increasing data diversity and thereby enhancing the model's generalization capability under limited sample conditions.

The dataset contains three insulator material categories: ceramics, composite, and glass. The image distribution across the three categories is 1,024 ceramic, 1,052 composite, and 1,013 glass images, totaling 3,089 images. A total of 7,119 insulator instances are annotated across all images, with an average of 2.30 insulator instances per image. The distribution of annotated instances is as follows: ceramics account for 3,666 instances, glass for 2,015 instances, and composite for 1,438 instances.

Figure 8
Diagram of Insulator Materials.



(a) ceramics (b) composite (c) glass

Annotation was performed using LabelImg software by two independent annotators. The annotation guideline required bounding boxes to tightly fit the outer edges of insulators. For partially occluded or blurred samples, annotation standards were determined through discussion between the two annotators. All annotation results were cross-validated, achieving over 95% consistency, with inconsistent samples being reviewed and corrected.

The dataset partitioning adopts a stratified sampling strategy to ensure that the proportion of each material type remains consistent across the training set, validation set, and test set, with a partitioning ratio of training set : validation set : test set = 7 : 2 : 1, providing a reliable data foundation for model training and evaluation.

3.3. Evaluation Indicators

In order to evaluate the efficiency of the FVS-YOLO model in detecting insulator materials, the average precision (AP), mean average precision (mAP), parameter quantity (Params), computational quantity (GFLOPs), and detection speed (FPS) indicators were used for assessment.

When calculating the AP value, an IoU threshold is usually set. When the IoU threshold is 0.5, the AP value is called the $AP_{0.5}$ average precision, which is the calculation of various targets, AP values. The calculation formula for AP is shown in Equation (7).

$$AP_i = \int_0^1 P(R)dR \,. \tag{7}$$

mAP as a commonly used comprehensive performance indicator, is typically used to evaluate the performance of multi-object detection algorithms. It is the mean average precision of all types of targets, and the calculation formula for mAP is shown in Equation (8).

$$mAP = \frac{1}{n}\left(\sum_{i=1}^{n} AP_i\right) \tag{8}$$

In the formula, AP_i represents the average accuracy of the i type of target, which is the area under the curve, P–R is the number of target categories, P is the precision, and R is the recall rate.

3.4. Ablation Experiment

To validate the effectiveness of each improved module and to determine whether conflicts exist among them, this study conducted a series of ablation experiments based on the YOLOv8s model, introducing each improvement incrementally to verify the effectiveness of the FVS-YOLO model. The experimental setup is as follows: Improvement 1 denotes replacing the backbone network with FasterNet; Improvement 2 denotes embedding a convolutional layer with a stride of 1 into the FasterNet backbone; Improvement 3 denotes replacing the neck C2f module with the VoVGSCSPC module; Improvement 4 denotes removing the large-scale detection layer from the network; Improvement 5 denotes embedding the SimAM attention mechanism before the detection head. All experiments were run five times independently (with random seeds set to 1, 2, 3, 4, and 5, respectively), and the results are reported as mean ± standard deviation. The experimental design and results are shown in Table 2, where "√" indicates the introduction of the corresponding improvement and "×" indicates its absence.

Table 2
Ablation experiment.

Area of improvement					mAP@50	Params(MB)	GFLOPs(G)	FPS(f/s)	Weight size(MB)
1	2	3	4	5					
×	×	×	×	×	0.916±0.012	11.1	28.4	194.8±2.3	21.4
√	×	×	×	×	0.901±0.015	5.6	14.0	241.6±3.1	10.8
√	√	×	×	×	0.912±0.011	7.3	18.3	217.6±2.7	14.1
√	√	√	×	×	0.907±0.013	6.8	16.2	214.0±2.5	13.3
√	√	×	×	√	0.913±0.010	7.3	18.3	218.4±2.8	14.1
√	√	√	√	×	0.911±0.014	6.8	16.2	231.3±2.9	13.3
√	√	√	√	√	0.916±0.090	5.0	15.1	229.9±2.6	9.8

Note: The mAP@50 and FPS values in the table are presented as mean ± standard deviation from five independent experiments. The standard deviation of all mAP@50 values is below 0.15%, indicating that the experimental results are stable and reliable. The parameters, GFLOPs, and model size are theoretical values and remained unchanged across the five experiments.

As shown in Table 2, replacing the backbone network of the YOLOv8s model with the lightweight FasterNet (Improvement 1) reduced the mAP@50 from 91.6% to 90.1%, a decrease of 1.5 percentage points. At the same time, the number of parameters decreased from 11.1M to 5.6M (a 49.5% reduction), computational cost dropped from 28.4 GFLOPs to 14.0 GFLOPs (a 50.7% reduction), and model size was compressed from 21.4 MB to 10.8 MB (a 49.5% reduction). This indicates that the FasterNet backbone effectively reduces both the time and space complexity of the model, albeit at the cost of some accuracy loss.

To mitigate the accuracy degradation caused by the insufficient feature extraction capability of the lightweight backbone, a convolutional layer with a stride of 1 was embedded into the FasterNet network (Improvement 2). As shown in Table 2, after incorporating Improvement 2, the mAP@50 increased from 90.1% to 91.2%, a gain of 1.1 percentage points. Although the parameter count, computational cost, and model size increased slightly (to 7.3M, 18.3 GFLOPs, and 14.1 MB, respectively), they remained substantially lower than those of the baseline YOLOv8s model. This demonstrates that embedding the convolutional layer effectively enhances the feature extraction capability of the network, thereby improving detection accuracy.

Replacing the C2f module in the neck network with the VoVGSCSPC module (Improvement 3) further reduced the model's parameter count, computational cost, and model size. As shown in the fourth row of Table 2, compared with the third row, the number of parameters decreased from 7.3M to 6.8M (a 6.8% reduction), computational cost dropped from 18.3 GFLOPs to 16.2 GFLOPs (an 11.5% reduction), and model size was reduced from 14.1 MB to 13.3 MB (a 5.7% reduction). This indicates that the VoVGSCSPC module achieves further lightweighting. It should be noted, however, that introducing this module caused a slight decrease in mAP@50 (from 91.2% to 90.7%), which needs to be compensated for in subsequent improvements.

After removing the large-scale detection layer from the neck network (Improvement 4), the model's parameter count decreased from 6.8M to 5.0M (a 26.5% reduction), computational cost dropped from 16.2 GFLOPs to 15.1 GFLOPs (a 6.8% reduction), and model size was compressed from 13.3 MB to 9.8 MB (a 26.3% reduction), while the mAP@50 increased from 90.7% to 91.1%. This suggests that removing the redundant large-scale detection layer not only reduces model complexity but also improves detection accuracy by reducing interference from irrelevant features.

After embedding the parameter-free SimAM attention mechanism before the detection head (Improvement 5), the complete model (FVS-YOLO) achieved an mAP@50 of 91.6%, matching the baseline YOLOv8s, while the number of parameters (5.0M), computational cost (15.1 GFLOPs), and model size (9.8 MB) were reduced by 54.95%, 46.83%, and 54.21%, respectively, compared to the baseline. The

detection speed increased from 194.8 FPS to 229.9 FPS, a gain of 18.0%. The ablation experiment results demonstrate that the various improved modules work well together, and the final model achieves significant lightweighting while maintaining detection accuracy.

3.5. Contrast Experiment

To comprehensively evaluate the performance of the proposed FVS-YOLO model, this study compared it with several mainstream object detection models, including YOLOv5s, YOLOv7tiny, YOLOv8s, YOLOv8n, YOLOv10s, YOLOv11n, YOLOv12n, YOLOv26n, as well as the lightweight model MobileNetV3-SSD. All models were trained and tested on the same dataset, and the experimental results are shown in Table 3.

As shown in Table 3, the FVS-YOLO model achieves an mAP@50 of 91.6%, matching that of YOLOv8s (also 91.6%), ranking among the top of the compared models. On the more stringent mAP@50:95 metric, FVS-YOLO attains 71.3%, slightly outperforming YOLOv8s (71.2%), indicating strong localization accuracy.

In terms of lightweight design, FVS-YOLO has 5.0M parameters, representing a 54.95% reduction compared to YOLOv8s; its computational cost is 15.1 GFLOPs, a 46.83% reduction; and its model size is 9.8 MB, a 54.21% reduction. Compared to YOLOv8n, which is specifically designed for lightweight applications, although YOLOv8n has fewer parameters (3.2M) and lower computational cost (8.7 GFLOPs), its mAP@50 is only 90.1%, which is 1.5 percentage points lower than that of FVS-YOLO. Compared to the ultra-lightweight YOLOv11n, YOLOv12n, and YOLOv26n, FVS-YOLO achieves 1.4, 2.4, and 3.9 percentage points higher in mAP@50, respectively, and 3.8, 3.9, and 5.7 percentage points higher in mAP@50:95, respectively, demonstrating a clear accuracy advantage. Compared to MobileNetV3-SSD, FVS-YOLO outperforms it by 3.2 percentage points in mAP@50, while the parameter count and model size are roughly comparable. This indicates that FVS-YOLO achieves a better balance between accuracy and efficiency.

In terms of detection speed, FVS-YOLO reaches 229.9 FPS, outperforming YOLOv8s (194.8 FPS) and YOLOv5s (192.3 FPS), and is slightly lower than YOLOv8n (267.5 FPS) as well as YOLOv11n (288.6 FPS), YOLOv12n (299.0 FPS), and YOLOv26n (328.5 FPS). Considering accuracy, speed, and model size comprehensively, FVS-YOLO demonstrates comprehensive advantages in the task of insulator material detection.

Regarding per-class AP, FVS-YOLO achieves the highest recognition accuracy on glass insulators (92.3%), followed by composite insulators (92.0%), and relatively lower accuracy on ceramic insulators (90.5%). The analysis suggests that glass insulators, with their smooth surfaces and pronounced reflective characteristics, are more distinguishable from

Table 3
Model Comparison Results.

Models	Ap (%)			Map@50 (%)	Map@50:90 (%)	Params (M)	GFLOPs (G)	FPS (f/s)	Weight size (MB)
	ceramic	composite	glass						
MobileNetV3-SSD	0.875	0.889	0.891	0.884	0.671	4.5	6.2	201.3	9.1
YOLOv5s	0.901	0.912	0.914	0.909	0.705	7.0	15.8	192.3	13.7
YOLOv7tiny	0.878	0.895	0.922	0.891	0.684	6.0	13.1	217.4	11.7
YOLOv8s	0.908	0.918	0.908	0.916	0.712	11.1	28.4	194.8	21.4
YOLOv8n	0.891	0.905	0.907	0.901	0.693	3.2	8.7	267.5	6.5
YOLOv10s	0.894	0.909	0.912	0.905	0.699	8.0	24.5	196.1	15.7
YOLOv11n	0.910	0.872	0.925	0.902	0.675	2.59	6.3	288.6	5.5
YOLOv12n	0.904	0.844	0.927	0.892	0.674	2.58	6.3	299.0	5.5
YOLOv26n	0.902	0.826	0.904	0.877	0.656	2.38	5.2	328.5	5.4
FVS-YOLO	0.905	0.920	0.923	0.916	0.713	5.0	15.1	229.9	9.8

the background, whereas ceramic insulators have rougher surfaces and colors similar to those of some composite insulators, increasing recognition difficulty. Compared to YOLOv7tiny, although its computational cost (13.1 GFLOPs) is slightly lower than that of FVS-YOLO, its mAP@50 is only 89.1%, the lowest among the compared models, making it insufficient for accurate insulator material recognition.

3.6. Results Visualization

The lightweight insulator recognition model based on YOLOv8s has significant improvements in terms of parameter quantity, computational load and weight size. To qualitatively evaluate the detection performance of the model, Figure 9 shows the visualization results of the lightweight insulator detection model. From the figure, it can be seen that the improved model in this paper can accurately identify insulators in power transmission lines. Therefore, the modern intelligent inspection method based on computer vision, compared with the traditional inspection method, demonstrates a significant advantage in inspection speed, and achieves a relative balance between inspection speed and inspection accuracy.

3.7. Robustness Analysis

To verify the robustness of the proposed model under complex environments, four typical interferences were respectively added to the test set: fog (density 0.02), low light (brightness 50%), Gaussian noise (σ = 0.05), and motion blur (7×7 kernel). The performance changes of FVS-YOLO and the baseline model YOLOv8s were tested under these conditions. The experimental results are presented in Table 4.

Table 4

Model performance comparison under different interference conditions (mAP@50).

Interference	YOLOv8s	FVS-YOLO
No interference	0.916	0.916
Fog	0.842	0.861
Low light	0.803	8.829
Gaussian noise	0.788	0.814
Motion blur	0.825	0.847

Figure 9

Visualized Results.



4. Conclusion

This study proposes FVS-YOLO, a lightweight insulator material detection model that focuses on small-scale features and integrates drone technology with deep learning to effectively identify insulator materials. In the FVS-YOLO model, FasterNet is employed to achieve lightweight design while maintaining stronger representational capacity through partial convolution, aiming to extract subtle features for material discrimination rather than merely localizing insulators. Furthermore, the VoVGSCSPC lightweight module is incorporated to further enhance model efficiency. Additionally, by removing large-scale detection layers and embedding the SimAM attention mechanism before the detection head, feature representation capability is strengthened to improve detection performance. These two improvements work synergistically to enhance the model's ability to focus on and fuse key material-specific features in complex backgrounds and multi-scale scenarios, thereby boosting the robustness and accuracy of material classification. An annotated dataset comprising various insulator materials has been created and is planned for public release, establishing the first performance benchmark for this specialized research direction, which holds significant importance for advancing the field.

Based on the comprehensive experimental analysis above, the results show that the parameter size of the FVS-YOLO model is 5.0M, the computational cost is 15.1G, and the weight size is 9.8MB, which are 54.95%, 46.83%, and 54.21% less than those of the YOLOv8s model, respectively. FVS-YOLO achieves a mAP@50 comparable to the state-of-the-art (0.916, on par with YOLOv8s). Experimental results demonstrate that the FVS-YOLO model achieves effective lightweight design while maintaining detection accuracy, with a 54.95% reduction in parameters, a 46.83% reduction in computational cost, and an 18.0% improvement in inference speed. Thus, the FVS-YOLO model ensures accuracy while significantly improving detection speed. The model is suitable for deployment on unmanned aerial vehicles, meets the real-time requirements of mobile platforms, and provides a feasible new method for insulator material detection.

References

1. Alcayde-García, F., Salmerón-Manzano, E., Montero, M. A., Alcayde, A., Manzano-Agugliaro, F. Power Transmission Lines: Worldwide Research Trends. Energies, 2022, 15(16), 5777. https://doi.org/10.3390/en15165777
2. Bharati, P., Pramanik, A. Deep Learning Techniques: R-CNN to Mask R-CNN: A Survey. Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2019, 2020, 657-668. https://doi.org/10.1007/978-981-13-9042-5_56
3. Bochkovskiy, A., Wang, C. Y., Liao, H. Y. M. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv Preprint arXiv:2004.10934, 2020.
4. Chen, G. C., Gu, G. M., Yu, X. N., Li, Z. B. Method for Defect Detection of Catenary Insulators by Combining the Features of PHOG and BOW-SURF. Journal of Chinese Mini-Micro Computer Systems, 2021, 42(1), 172-177.
5. Fang, G., An, X., Fang, Q., Gao, S. Insulator Defect Detection Based on YOLOv5s-KE. Electronics, 2024, 13(17), 3483. https://doi.org/10.3390/electronics13173483
6. Girshick, R., Donahue, J., Darrell, T., Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, 580-587. https://doi.org/10.1109/CVPR.2014.81
7. Han, J., Wang, Z., Ru, S., Yang, W., Xu, X. Innovative Research on Multispectral Image Fusion Technology for Defect Detection in Composite Insulators of Ultra-High Voltage Transmission Lines. Australian Journal of Electrical and Electronics Engineering, 2025, 1-27. https://doi.org/10.1080/1448837X.2025.2553961
8. Hui, X., Bian, J., Zhao, X., Tan, M. Vision-Based Autonomous Navigation Approach for Unmanned Aerial Vehicle Transmission-Line Inspection. International Journal of Advanced Robotic Systems, 2018, 15(1), 1729881417752821. https://doi.org/10.1177/1729881417752821
9. Jenssen, R., Roverso, D. Automatic Autonomous Vision-Based Power Line Inspection: A Review of Current Status and the Potential Role of Deep Learning. International Journal of Electrical Power & Energy

Systems, 2018, 99, 107-120. https://doi.org/10.1016/j.ijepes.2017.12.016

10. Kaur, R., Singh, S. A Comprehensive Review of Object Detection with Deep Learning. Digital Signal Processing, 2023, 132, 103812. https://doi.org/10.1016/j.dsp.2022.103812
11. Li, J., Zhou, H., Lv, G., Chen, J. A2MADA-YOLO: Attention Alignment Multiscale Adversarial Domain Adaptation YOLO for Insulator Defect Detection in Generalized Foggy Scenario. IEEE Transactions on Instrumentation and Measurement, 2025, 74, 1-19. https://doi.org/10.1109/TIM.2025.3541814
12. Li, Y., Hu, B., Wei, S., Gao, D., Shuang, F. Simplify-YOLOv5m: A Simplified High-Speed Insulator Detection Algorithm for UAV Images. IEEE Transactions on Instrumentation and Measurement, 2025, 74, 3519614. https://doi.org/10.1109/TIM.2025.3549909
13. Li, Y., Li, L., Zhang, X., et al. Review of New Non-Destructive Testing Methods for Insulators Based on Sound, Light and Heat Characteristics. Insulation Materials, 2023, 56(10), 12-21.
14. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A. C. SSD: Single Shot Multibox Detector. European Conference on Computer Vision, 2016, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2
15. Liu, Y., Guo, Y., Wu, G., Fan, Y., Xiao, S., Zhang, X. A New Detection Method for the Ageing State of Composite Insulators Based on Spectral-Spatial Feature Fusion. Expert Systems with Applications, 2024, 254, 124353. https://doi.org/10.1016/j.eswa.2024.124353
16. Ma, L., Cui, W., Yu, D. Analysis of Three Insulator Materials and Their Research Progress. New Chemical Materials, 2011, 39(12), 28-30.
17. Mendu, B., Mbuli, N. State-of-the-Art Review on the Application of Unmanned Aerial Vehicles (UAVs) in Power Line Inspections: Current Innovations, Trends, and Future Prospects. Drones, 2025, 9(4), 265. https://doi.org/10.3390/drones9040265
18. Nooralishahi, P., Ramos, G., Pozzer, S., Ibarra-Castanedo, C., Lopez, F., Maldague, X. P. V. Texture Analysis to Enhance Drone-Based Multi-Modal Inspection of Structures. Drones, 2022, 6(12), 407. https://doi.org/10.3390/drones6120407
19. Redmon, J., Divvala, S., Girshick, R., Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, 779-788. https://doi.org/10.1109/CVPR.2016.91
20. Redmon, J., Farhadi, A. YOLOv3: An Incremental Improvement. arXiv Preprint arXiv:1804.02767, 2018.
21. Ren, S., He, K., Girshick, R., Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6), 1137-1149. https://doi.org/10.1109/TPAMI.2016.2577031
22. Wang, B., Ren, M., Xia, C., Li, Q., Dong, M., Zhang, C., Guo, C., Liu, W., Pischler, O. Evaluation of Insulator Aging Status Based on Multispectral Imaging Optimized by Hyperspectral Analysis. Measurement, 2022, 205, 112058. https://doi.org/10.1016/j.measurement.2022.112058
23. Yang, C. Insulator Detection Method Based on Texture Features. Electrical Technology, 2010, (7), 46-48, 52.
24. Zaidi, S. S. A., Ansari, M. S., Aslam, A., Kanwal, N., Asghar, M., Lee, B. A Survey of Modern Deep Learning Based Object Detection Models. Digital Signal Processing, 2022, 126, 103514. https://doi.org/10.1016/j.dsp.2022.103514
25. Zhang, Q., Zhang, J., Li, Y., Zhu, C., Wang, G. ID-YOLO: A Multi-Module Optimized Algorithm for Insulator Defect Detection in Power Transmission Lines. IEEE Transactions on Instrumentation and Measurement, 2025, 74, 3505611. https://doi.org/10.1109/TIM.2025.3527530

This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 (CC BY 4.0) License (http://creativecommons.org/licenses/by/4.0/).