energies

MDPI

Article

An Enhanced Lightweight YOLOv11 Algorithm for Real-Time Detection of High-Voltage Line Insulators

Abdil Karakan

Dazkırı Vocational School, Afyon Kocatepe University, 03950 Afyonkarahisar, Türkiye; akarakan@aku.edu.tr

Abstract

UAV-based insulator detection is challenging because insulators often occupy small regions of aerial images and appear against complex backgrounds, while subtle local features may be lost during feature extraction and down-sampling. Moreover, practical UAV and edge-device applications require efficient models with limited computational and memory demands. This study proposes an optimized lightweight YOLOv11n model for high-voltage transmission-line insulator detection. The architecture integrates C3k2MBNV2 to reduce model complexity, SCDown to preserve spatial information during down-sampling, and C3k2WTDC to enhance multi-frequency feature representation. A diverse dataset containing 5750 insulator images acquired under different environmental conditions, viewing angles, and backgrounds was used for evaluation. Experimental results show that the proposed model reduces the parameter count from 6.20 M to 3.26 M and computational complexity from 20.5 to 12.7 GFLOPs, corresponding to reductions of 47.4% and 38.0%, respectively. Meanwhile, precision increases from 91.3% to 93.8%, recall from 73.4% to 75.2%, mAP50 from 71.2% to 73.9%, and mAP50–95 from 65.6% to 67.3%. These results demonstrate an improved accuracy–efficiency trade-off, supporting real-time insulator detection in resource-constrained UAV and edge-device applications.

Keywords: YOLOv11; high-voltage insulator detection; lightweight deep learning; UAV image analysis; real-time object detection

1. Introduction

Insulators used in high-voltage transmission lines serve the dual purpose of providing electrical insulation and mechanically supporting conductors. Therefore, reliable detection of insulator defects is important for maintaining the safe operation of transmission-line infrastructure [1]. During long-term operation, insulators may be affected by environmental and mechanical stresses that can lead to surface damage, cracking, flashover, and other defects, potentially compromising the reliability of transmission systems [2]. Traditional manual inspection of transmission-line insulators is labour-intensive, time-consuming, and potentially hazardous, particularly in geographically challenging or high-voltage environments [3].

Conventional inspection methods typically involve manual inspection by field personnel, infrared imaging, or helicopter-based monitoring. These approaches face several limitations, including high costs, low efficiency, susceptibility to human error, and difficulties in accessing wide or remote areas. Consequently, more efficient and automated inspection approaches have received increasing attention. UAV-based transmission-line inspection provides a practical means of acquiring visual data from power-line infrastructure [4]. UAV

Check for updates

Academic Editor: Krzysztof Górecki

Received: 13 July 2026
Revised: 14 August 2026
Accepted: 20 August 2026
Published: 21 August 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

imaging has also become an effective approach for acquiring visual information for the automatic detection of outdoor high-voltage insulators [5]. However, the large volume of visual data generated during automated power-line inspection creates a need for efficient detection algorithms. Recent studies have therefore investigated lightweight deep-learning architectures, including improved YOLOv8n-based models, to reduce computational complexity while maintaining effective power-line defect detection [6].

Deep-learning-based methods have significantly advanced automatic object detection by reducing dependence on manually designed features. Convolutional Neural Networks (CNNs) can learn hierarchical representations ranging from low-level visual patterns to increasingly abstract semantic features. Among widely used CNN architectures, ResNet introduced residual connections to facilitate the training of deeper neural networks and improve feature representation [7]. For real-time object detection, one-stage detectors such as the YOLO family provide an efficient alternative to conventional two-stage detection frameworks. YOLOv3 further improved the YOLO framework by performing predictions at multiple scales, thereby enhancing its capability to detect objects of different sizes [8]. In general, deep-learning-based object detectors involve different trade-offs between detection performance and computational efficiency [9].

Despite these advances, UAV-based insulator detection presents several task-specific challenges. Insulators may occupy relatively small regions of aerial images because of variations in flight altitude, camera-to-object distance, and viewing angle. Local defects such as cracks, fractures, and surface anomalies are often represented by subtle visual details that may be weakened or lost during feature extraction and down-sampling. Transmission-line scenes may also contain complex backgrounds and visually interfering structures, increasing the likelihood of detection errors. Furthermore, the limited availability of sufficiently labelled insulator images can restrict model generalization. These challenges become particularly important in practical UAV and edge-based inspection systems, where computational complexity, memory consumption, and inference latency must also be considered. Therefore, an effective insulator detector should simultaneously preserve fine spatial information, represent subtle local characteristics, and maintain a lightweight architecture suitable for resource-constrained deployment.

The YOLO family has been widely adopted for real-time object detection because of its favourable balance between detection performance and computational efficiency. Earlier architectures, including YOLOv3 and YOLOv4, established efficient one-stage detection frameworks with strong real-time capabilities [10]. More recent studies have adapted newer YOLO architectures specifically to insulator inspection. Qin et al. developed an improved YOLOv8-based method for insulator defect detection in complex environments [11], while You and Zhao proposed an improved YOLOv7-based network specifically for insulator defect detection in UAV aerial images [12]. Recent studies have further focused on improving detection performance and computational efficiency through specialized lightweight architectures. Jiang et al. proposed the CDDCR-YOLOv8 algorithm for insulator defect detection [13], while Wang et al. developed an improved YOLOv7 model to enhance insulator defect recognition [14]. Han et al. introduced LMD-YOLO, a lightweight YOLOv8-based architecture designed for multi-defect detection of power-distribution-network insulators [15]. More recently, Zhao et al. proposed an improved YOLOv11n-based insulator defect detector incorporating architectural modifications for enhanced detection performance and lightweight deployment [16]. These studies demonstrate the effectiveness of recent YOLO architectures for insulator inspection; nevertheless, simultaneously improving detection performance, reducing computational complexity, and preserving fine-scale spatial information remains challenging, particularly for UAV-based inspection and resource-constrained deployment.

In addition to YOLO-based approaches, alternative architectures have been investigated for insulator inspection. Dian et al. proposed Faster R-Transformer for insulator detection in complex aerial environments, integrating Transformer-based feature representation into a two-stage detection framework [17]. Hien et al. investigated a different strategy by combining an improved Feature Pyramid Network (FPN), a lightweight GELAN architecture, and a diffusion model for insulator defect detection [18]. These studies demonstrate the increasing diversity of feature-extraction and feature-fusion strategies for insulator inspection. However, simultaneously achieving lightweight computation, preservation of fine spatial information, and robust detection in complex UAV scenes remains an open challenge.

YOLO11 introduces architectural refinements in its backbone and neck to improve feature extraction while maintaining computational efficiency, making it suitable for real-time object detection and deployment across different hardware environments [19]. Nevertheless, the baseline YOLO11n architecture is not specifically designed to address the combined challenges of small-target representation, fine-detail preservation, and computational constraints encountered in UAV-based insulator inspection.

Despite the progress achieved by existing lightweight and YOLO-based insulator detectors, several limitations remain. Many existing approaches primarily focus on improving detection accuracy through increasingly complex feature extraction or fusion mechanisms, whereas computational efficiency, preservation of fine spatial information, and effective representation of local insulator characteristics are not always addressed simultaneously. This trade-off becomes particularly important in UAV-based inspection, where insulators may appear as small targets in complex backgrounds and the detection model may ultimately need to operate under limited computational and memory resources. Therefore, there remains a need for an architecture that jointly considers lightweight computation, spatial-detail preservation, and effective multi-scale/multi-frequency feature representation within a unified detection framework.

To address this gap, this study develops a lightweight YOLOv11n-based architecture in which three complementary modifications are integrated according to the requirements of UAV-based insulator detection. Rather than optimizing a single performance dimension, the proposed design aims to simultaneously reduce model complexity, preserve spatial information during down-sampling, and enhance the representation of local and multi-frequency features.

To address these limitations in the baseline YOLOv11n architecture, the proposed model introduces three complementary architectural modifications, each targeting a specific constraint of UAV-based insulator detection. First, C3k2MBNV2 is employed to reduce parameter redundancy and computational cost through lightweight inverted residual and linear bottleneck operations. Second, SCDown is introduced to mitigate the loss of fine spatial information during down-sampling, which is particularly important for small and distant insulators occupying limited image regions. Third, C3k2WTDC incorporates wavelet-based convolution to complement conventional spatial-domain feature extraction with multi-frequency representations, allowing coarse structural information and fine local variations to be processed more effectively. Rather than treating these components as isolated substitutions, the proposed architecture integrates them as complementary mechanisms targeting model complexity, spatial-detail preservation, and frequency-aware feature representation, respectively.

The main contributions of this study can be summarized as follows:

A lightweight C3k2MBNV2 block is integrated into the YOLOv11n backbone to reduce parameter redundancy and computational complexity while maintaining effective feature representation for insulator detection.

The SCDown module is incorporated into the backbone and neck to reduce information loss during down-sampling and improve the preservation of spatial details associated with small and distant insulators.

A C3k2WTDC block based on wavelet-transform depth-wise convolution is introduced into the neck to enhance multi-frequency feature representation while maintaining a lightweight computational structure.

The three complementary modifications are integrated into a unified lightweight YOLOv11n framework, and their effects on detection performance and computational efficiency are evaluated through comparative and ablation experiments.

2. Materials and Methods

The insulator detection system developed in this study begins with the pre-processing of high-resolution images acquired via UAVs and proceeds with an optimized deep learning architecture based on YOLOv11n. Initially, the input images are normalized at the input layer and transformed into a format suitable for the model. Subsequently, the C3k2MBNV2 blocks located in the backbone extract essential spatial and semantic features. During this process, the SCDown module operates by decoupling spatial and channel information to minimize information loss during down sampling. In the neck section, the C3k2WTDC block integrates multi-scale frequency analysis through wavelet transform, enabling the precise extraction of complex insulator features. Finally, the detection head generates bounding boxes and class predictions. This architecture results in a real-time, high-accuracy detection system with low resource consumption, suitable for field deployment. The overall workflow of the proposed insulator detection system is illustrated in Figure 1.

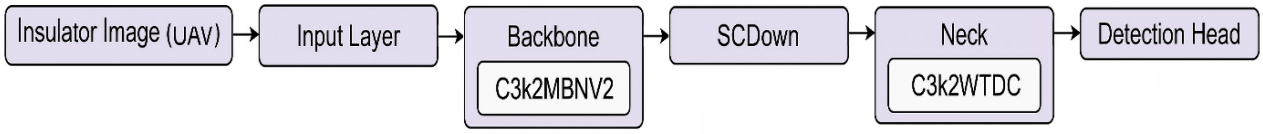


Figure 1. General workflow of the proposed isolator detection system.

2.1. Data Set Preparation

To enhance the effectiveness of the proposed lightweight YOLOv11 architecture in the real-time and accurate detection of high-voltage line insulators, a high-quality, specifically structured training dataset was created. Since the performance of deep learning-based detection models heavily depends on the diversity and representativeness of the training data, various insulator images captured under different environments, angles, and conditions were included to improve the model's generalization capability.

The dataset comprises images of both healthy insulators and those exhibiting potential defects such as cracks, fractures, and surface contamination. This diversity was intentionally designed to enable the model to reliably distinguish between intact and faulty insulators. A significant portion of the images were captured in high resolution using UAVs, with careful consideration of varying altitudes and viewpoints. Moreover, images were collected under direct sunlight as well as in shaded environments, ensuring consistent model performance under challenging environmental conditions.

Video footage was segmented into frames, and only high-quality, clear, and information-rich images were selected. These selected frames were manually annotated, with each insulator labelled with class tags and bounding box coordinates in a format compatible with the YOLOv11 algorithm.

The insulator images in the dataset were obtained from different infrastructures (steel lattice towers, concrete poles) and geographical locations. Furthermore, the number and arrangement of insulators in the images varied, allowing the model to perform accurately even in complex scenes involving multiple insulators. During the annotation process, careful attention was paid to the complete and precise labelling of insulators to minimize deviations during training.

Another key aspect was the inclusion of various scenarios in the dataset that preserve the visual distinguishability of insulators despite the presence of visual noise, complex environmental backgrounds, and low-contrast conditions. This approach aimed to ensure that the developed model performs with high accuracy in real-time applications and remains robust against diverse field conditions.

The comprehensive and balanced dataset constructed in this study provided a solid foundation for training the YOLOv11-based insulator detection system, facilitating the optimization of both detection accuracy and inference speed. Figure 2 presents the number of sample images representing different types and characteristics in the UAV-based insulator image dataset.

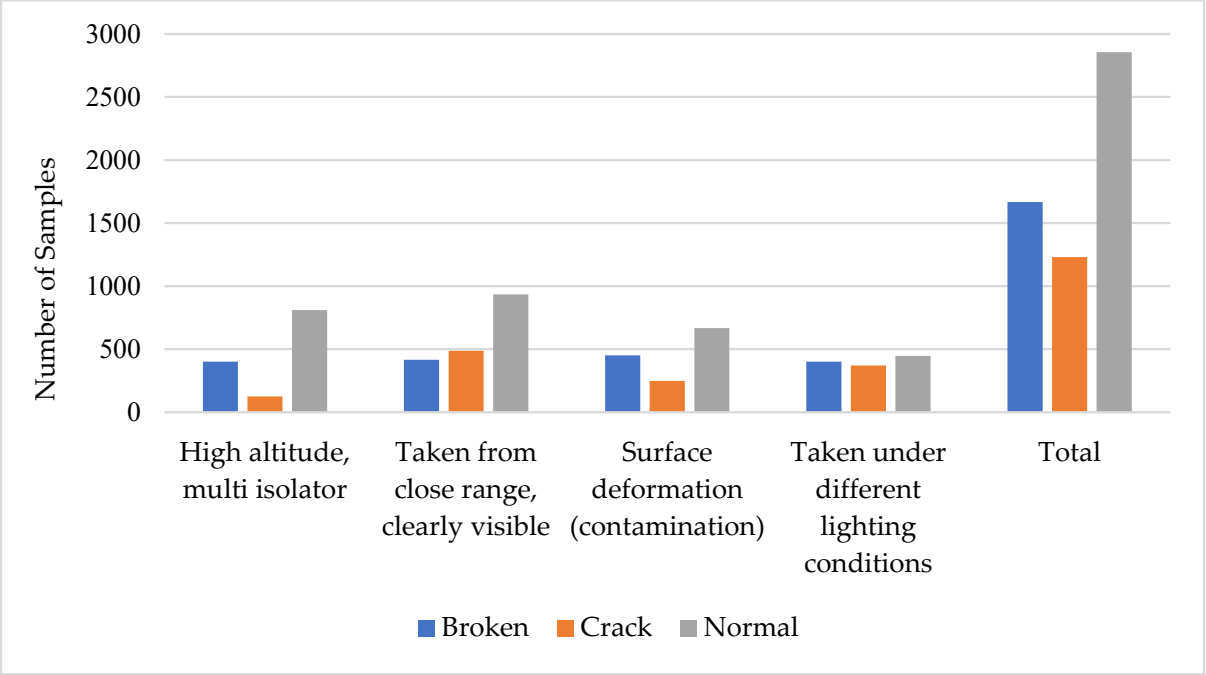


Figure 2. Sampling numbers of different types and features used in the UAV-based isolator image dataset.

The success of the developed insulator detection system is largely dependent on the quality and diversity of the constructed dataset. By incorporating images collected under various lighting and weather conditions from different geographical regions, the generalization capability of the model has been significantly enhanced. A balanced representation of both healthy and defective insulators has been ensured within the dataset. This comprehensive and meticulously annotated dataset enables the model to perform with high accuracy and stability in real-world scenarios, thereby directly reinforcing the practical applicability of the system in field deployments. Figure 3 illustrates sample insulator images obtained under different conditions.



Figure 3. Representative isolator images obtained under different conditions.

2.2. YOLOv11

Among real-time object detection algorithms, the YOLO (You Only Look Once) family stands out due to its ability to combine high accuracy with low latency, making it suitable for a wide range of applications [20]. The latest version in the series, YOLOv11, comprises four main architectural components: the input layer, the backbone, the neck, and the detection head. Together, these components enable the model to detect objects at multiple scales while operating efficiently with limited hardware resources.

The backbone of YOLOv11 is designed to extract features in a compact and effective manner using the C3k2 block as its core unit. This block, which includes dual convolutional layers with a kernel size of 2, facilitates the extraction of deep features with reduced computational demand. Multi-scale feature integration is achieved via the Spatial Pyramid Pooling—Fast (SPP-Fast) module, which transforms spatial information from varying receptive fields into a unified representation, thereby enhancing the detection of targets with diverse dimensions.

To better process spatially significant regions, the model's neck incorporates a Convolutional Block with Parallel Spatial Attention (C2PSA). This module leverages a parallel spatial attention mechanism to focus on critical regions within the image, significantly improving detection accuracy. Additionally, the dual-path feature aggregation strategy based on the C3k2 block facilitates effective interaction between low-level and high-level features.

The CBS block used in the up-sampling phase enhances feature sharpness, while the cross-stage connections employed in the down sampling stage strengthen information flow between deep and shallow layers, thereby improving the model's discriminative capability. Moreover, the binary Depthwise Convolution (DWConv) structure in the classification layer reduces unnecessary computations, decreasing both the number of parameters and the overall processing cost.

YOLOv11 is offered in multiple model sizes, all sharing a common architecture, to accommodate object detection tasks at varying scales. In applications such as high-voltage insulator detection, it is crucial that the model not only achieves high accuracy but also performs real-time inference with minimal latency. Therefore, minimizing the computational load is essential for deployment on resource-constrained edge devices.

In this study, the core architecture of YOLOv11 was redesigned, particularly by optimizing the backbone and neck sections to be more lightweight and efficient. This optimization led to significant reductions in the total number of parameters and FLOPs, while simultaneously improving accuracy, resulting in a model architecture well-suited for real-time applications. The overall architecture of the proposed model is illustrated in Figure 4.

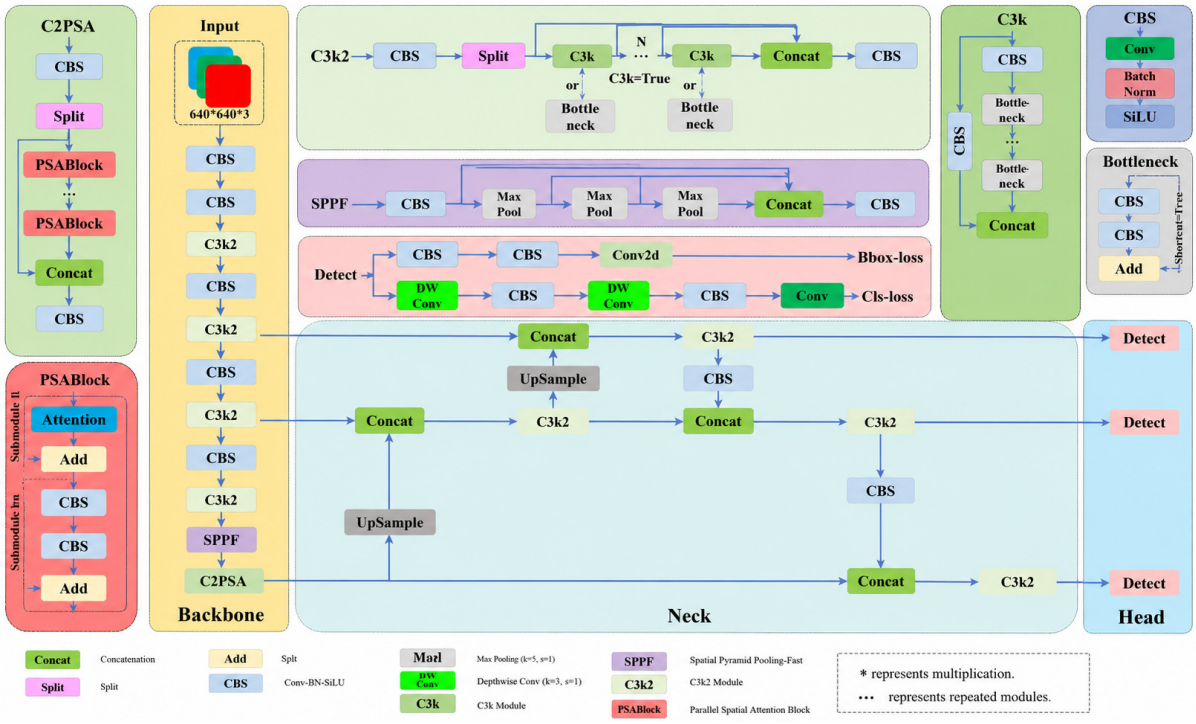


Figure 4. Structural Architecture of the Algorithm [21].

2.3. Lightweight YOLOv11 Model Design

2.3.1. Development of C3k2MBNV2 Block Structure

In the backbone architecture of YOLOv11, the C3k2 block has been integrated to enhance the multi-scale feature extraction capability. This structure is based on the Cross-Stage Partial (CSP) mechanism and utilizes low-dimensional kernels to generate denser feature maps. Particularly in complex scenes such as insulator detection, this block provides detailed spatial information that enables the precise perception of visual elements. However, this architecture also introduces certain limitations, including increased computational load and higher hardware requirements.

To address this issue, a more lightweight and efficient structure, named the C3k2MBNV2 block, was designed by drawing inspiration from the MobileNetV2 architecture. This block employs inverted residual connections and linear bottleneck structures to significantly reduce both the FLOPs and the number of parameters, without compromising the feature representation capacity [22].

Although more recent lightweight backbones such as MobileViT, GhostNet, and EfficientNet-Lite are available, MobileNetV2 was selected in this study because its inverted residual and linear bottleneck operations can be integrated into the existing C3k2/CSP struc-

ture with relatively limited architectural modification. Its depth-wise separable convolution mechanism directly targets the parameter and computational redundancy that the proposed C3k2MBNV2 block is intended to reduce. In contrast, architectures such as MobileViT introduce additional transformer-based operations, while GhostNet and EfficientNet-Lite employ different feature-generation and scaling strategies that would require more extensive restructuring of the YOLOv11n backbone. Therefore, the choice of MobileNetV2 was based not solely on architectural recency, but on its structural compatibility with the C3k2 framework, implementation simplicity, and favourable efficiency–representation trade-off for the targeted lightweight redesign.

Unlike conventional residual connections, the inverted residual block begins by expanding the channel dimensions using a 1×1 pointwise convolution. Then, a 5×5 depth-wise convolution is applied to extract spatial features. Finally, a linear projection is used to reduce the channel size again. This structure reduces information loss while enhancing computational efficiency. Notably, the use of linear activation functions in the dimensionality reduction stage helps prevent representational degradation commonly associated with traditional non-linear activations.

In addition, the developed architecture incorporates the Cross-Bottleneck (CB) module. This module employs a 1×1 convolution kernel and removes the SiLU activation function, which reduces the number of channels and consequently the computational load, while maintaining residual connections to strengthen information flow within the network. As a result, the network can extract high-level features efficiently and transmitting more information with fewer resources.

As a result of these architectural improvements, the traditional bottleneck blocks in the YOLOv11 backbone have been replaced with MBNV2 blocks. This modification is intended to reduce model complexity while maintaining the feature representation required for small and distant insulators under complex scene conditions. The combination of all these improvements increases the overall accuracy of the model while optimizing resource consumption. The architectural structure of the proposed C3k2MBNV2 block is illustrated in Figure 5.

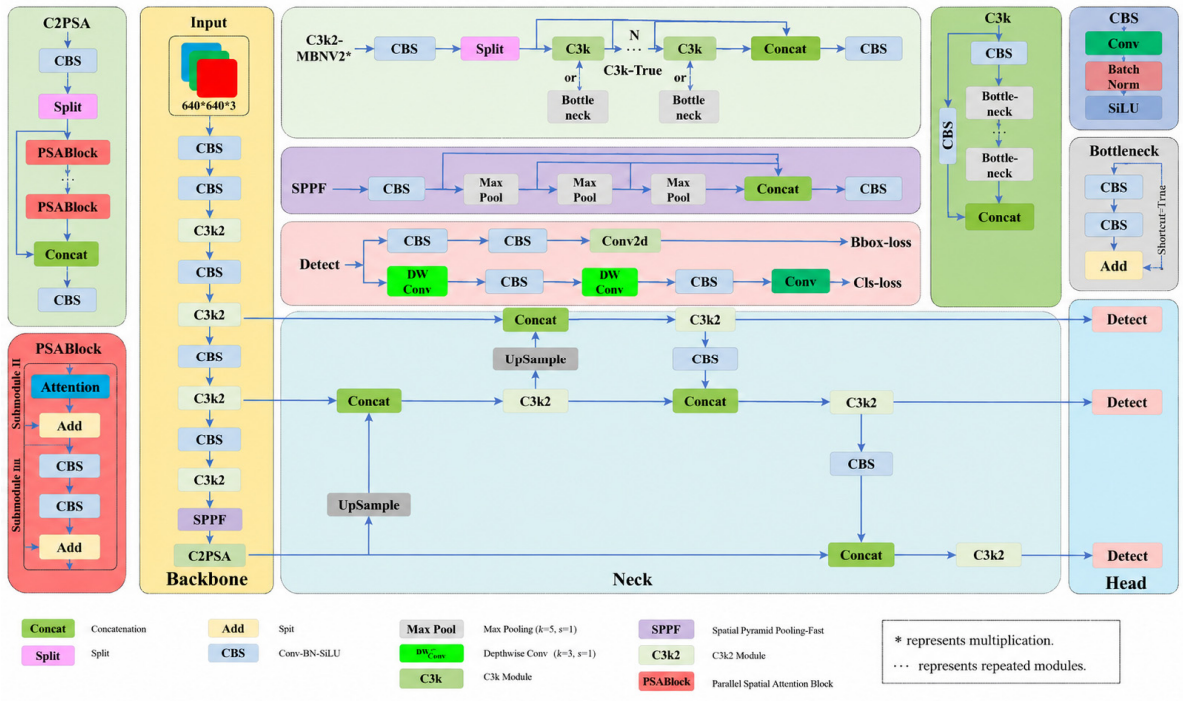


Figure 5. Improved C3k2MBNV2 Backbone Block Architecture [21].

2.3.2. Multi-Scale Feature Aggregation with SCDown Block Integration

Although CNNs generate feature maps with rich semantic information in deeper layers, their low spatial resolution can hinder accurate detection of small objects. In applications such as high-voltage insulator detection where objects are relatively small effectively integrating features from both shallow and deep layers becomes crucial.

To address this challenge, the SCDown block was integrated into both the backbone and neck components of the YOLOv11 model [23]. This block is designed to minimize the information loss commonly caused by traditional down sampling methods, while maintaining computational efficiency. In the first stage, SCDown transforms the channel dimensions of the input feature map via pointwise convolution (1×1), without altering the spatial resolution. Next, a depth-wise separable convolution is applied, performing sampling only in the spatial domain while preserving channel-specific information. As a result, the output maintains the original number of channels, despite the reduced spatial resolution.

Through this architecture, the spatial information of each channel is processed independently, effectively mitigating the loss of fine details often observed in standard convolutional operations. This ensures the retention of crucial features associated with small objects. Moreover, the design supports multi-scale feature fusion, allowing the model to leverage both shallow and deep representations in a unified manner.

Another key advantage of the SCDown block lies in its reduced computational cost. Unlike standard convolutions, which operate across all channels simultaneously, SCDown relies on channel-wise operations, leading to significantly lower FLOPs. This efficiency is particularly beneficial for edge devices, where computational resources are limited.

The relevance of SCDown to small-object insulator detection arises from its decoupled spatial and channel processing strategy. Conventional stride-based down-sampling simultaneously reduces spatial resolution and transforms channel information, which may cause fine object boundaries and localized responses to be weakened, particularly when the target occupies only a small region of the feature map. In SCDown, spatial reduction and channel-wise feature processing are handled more selectively, allowing informative local responses to be retained with lower computational redundancy. This characteristic is particularly relevant for distant insulators and small defect regions in UAV imagery, where excessive spatial-information loss at early or intermediate network stages can negatively affect subsequent detection. The overall architecture integrating the SCDown block is illustrated in Figure 6.

2.3.3. Improving Multi-Scale Feature Extraction with C3k2WTDC Block Design

Real-time and highly accurate detection of high-voltage power line insulators requires the effective extraction of fine details embedded in images across multiple frequency domains. In this context, local defects such as cracks or fractures on insulator surfaces represent high-frequency features confined to small regions. In contrast, defects such as surface contamination or colour irregularities are distributed over broader areas and correspond to lower-frequency content. Traditional convolutional architectures, due to their limited receptive fields, struggle to adequately represent such multi-scale variations. Moreover, aliasing effects caused by standard down sampling operations lead to the loss of high-frequency information, thereby negatively impacting detection accuracy. Although large-kernel convolutions can increase the receptive field, they also significantly inflate the number of model parameters, making them inefficient for deployment on resource-constrained embedded systems [24].

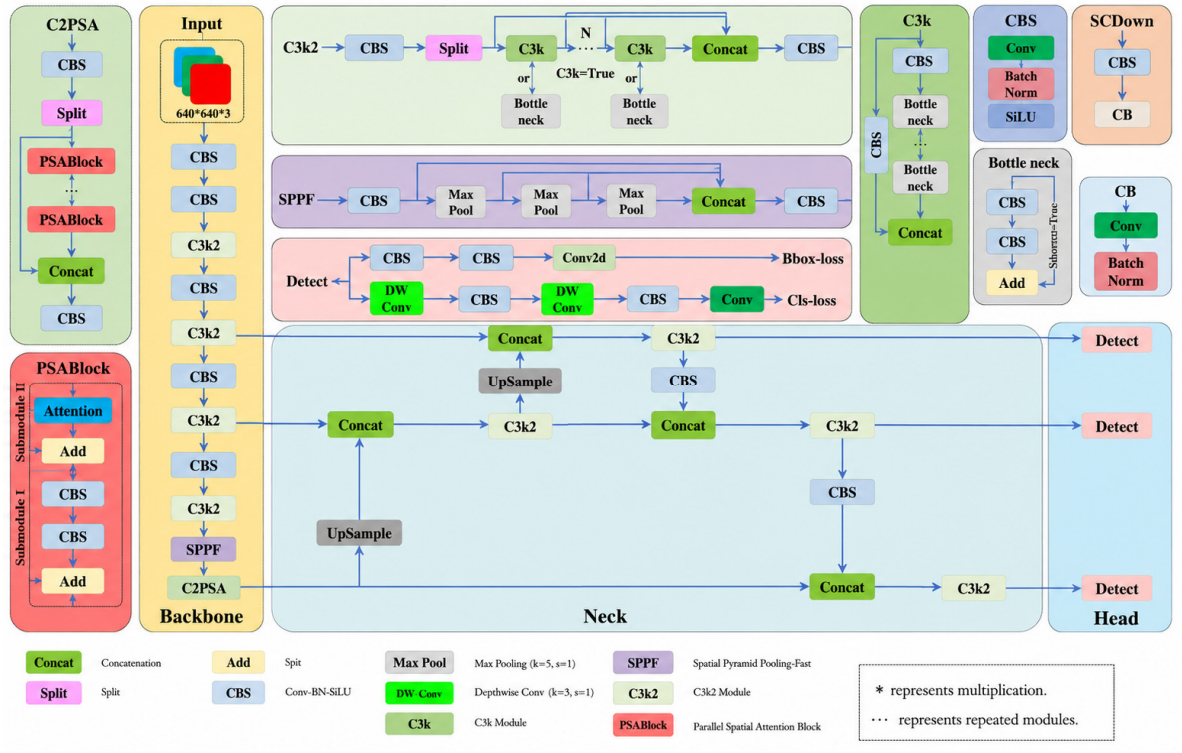


Figure 6. SCDown Module Structure [21].

To overcome these limitations, a wavelet transform-based convolutional block (C3k2WTDC) was integrated into the neck structure of the proposed model. Within this block, the Wavelet Transform Convolution (WTConv) layer decomposes the input feature map into multiple frequency bands and applies separate convolutional operations to each component. This enables effective extraction of both low- and high-frequency insulator features while also achieving a large receptive field using small convolutional kernels.

Following WTConv, the WTPC (Wavelet Transform Pointwise Convolution) layer utilizes 1×1 convolutions to reorganize channel dimensions, preserving spatial detail while reducing computational complexity. This approach reduces the model's overall FLOPs and facilitates the integration of deep feature representations.

In the final stage, the WTDC (Wavelet Transform Depth-wise Convolution) layer independently processes each channel. By performing per-channel operations, this depth-wise convolution prevents unnecessary parameter growth and enhances the model's ability to distinguish subtle surface anomalies on the insulator.

The combined architecture of the C3k2WTDC block strengthens feature extraction through multi-scale frequency analysis while keeping the model's parameter count and computational load minimal. Consequently, the proposed lightweight YOLOv11 architecture is optimized for high-precision, low-latency detection on edge devices. The incorporation of this block significantly improves the detection performance of small and irregularly shaped objects, such as insulators, even under challenging environmental conditions.

From a frequency-domain perspective, the wavelet decomposition separates the input feature maps into low-frequency (LL) and high-frequency directional components (LH, HL, and HH). The LL component mainly retains coarse structural and contextual information, whereas the high-frequency components emphasize rapid spatial variations, including edges, boundaries, and fine local patterns. These characteristics are particularly relevant to insulator inspection because small cracks, fractures, and boundary irregularities may produce localized intensity transitions that can be weakened during conventional convolution and repeated down-sampling. By processing these frequency components separately

before feature reconstruction, C3k2WTDC provides a mechanism for retaining both global structural information and fine local variations. Therefore, the role of the wavelet operation in the proposed architecture is not merely dimensional transformation; rather, it provides complementary frequency-domain information that may assist the detector in representing subtle insulator structures. To further illustrate the mechanism of the proposed C3k2WTDC block, Figure 7 presents the wavelet-based multi-frequency feature extraction process. The input feature representation is decomposed into the LL, LH, HL, and HH components, enabling coarse structural information and fine directional details to be processed separately before feature fusion.

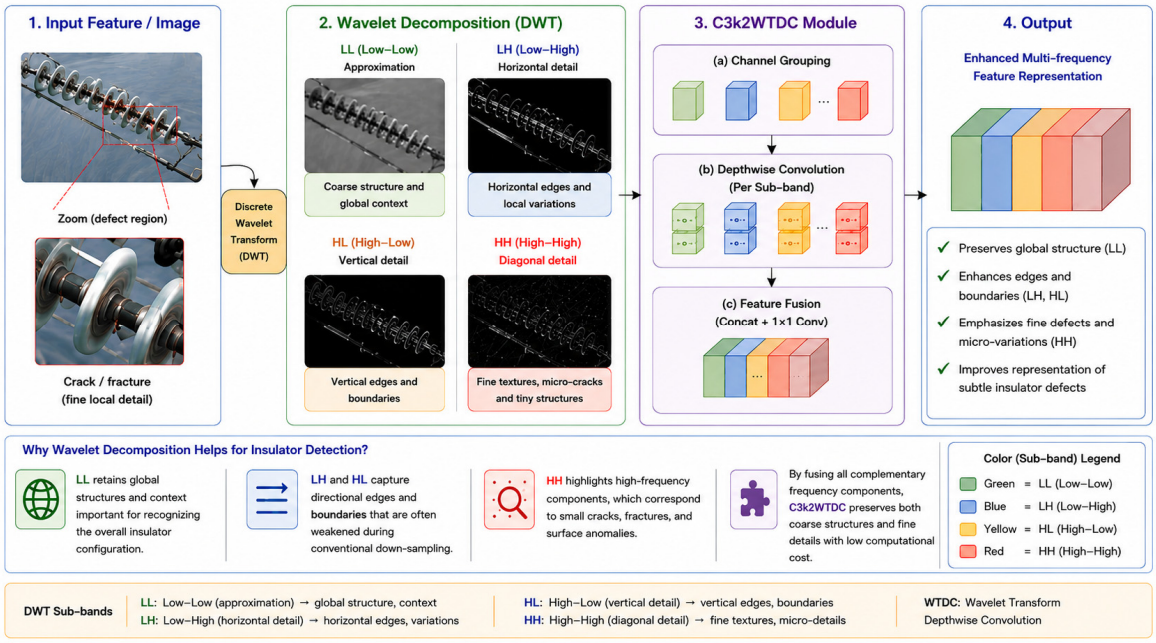


Figure 7. Mechanistic illustration of wavelet-based multi-frequency feature extraction in the C3k2WTDC module.

Figure 7 therefore provides a mechanistic interpretation of how C3k2WTDC complements conventional spatial-domain feature extraction. By jointly exploiting structural and high-frequency detail information, the module is designed to improve the representation of subtle insulator characteristics without relying on computationally expensive large-kernel operations.

2.3.4. Improved Lightweight YOLOv11n Model

The proposed model architecture enhances computational efficiency and ensures high accuracy in feature extraction through three key structural improvements. These enhancements include: the C3k2MBNV2 block integrated into the backbone structure, the SCDown module positioned within both the backbone and neck parts, and the C3k2WTDC block located in the neck section.

Firstly, the C3k2MBNV2 block presents a lightweight structure equipped with inverted residual connections and a linear bottleneck design, enabling the preservation of critical features related to insulators while minimizing unnecessary parameters. This structure offers high performance with lower computational cost, particularly suitable for resource-constrained platforms.

Secondly, the SCDown module, applied to both the backbone and neck layers, processes spatial and channel information separately. This reduces aliasing effects during the

down sampling process and helps preserve fine details in complex scenes. As a result, it enables more successful detection of small-sized insulator structures.

The third improvement, the C3k2WTDC block, incorporates wavelet transform-based convolutional layers (WTConv) combined with multiple convolution operations, thereby enabling multi-frequency information processing. This facilitates effective analysis of both small- and large-scale insulator features, while maintaining a small kernel size to keep the computational load minimal.

The selection and integration of these three modules were guided by complementary design objectives rather than by independent module substitution. C3k2MBNV2 primarily addresses the computational redundancy of the baseline backbone by employing inverted residual and depth-wise convolution operations to reduce parameter and computational requirements while maintaining feature representation capability. In contrast, SCDown addresses the spatial-information loss associated with conventional down-sampling by decoupling spatial and channel-wise processing, which is particularly relevant for small and distant insulators in UAV imagery. C3k2WTDC complements these components by incorporating wavelet-based depth-wise convolution to represent information at different frequency components while maintaining computational efficiency. Consequently, the three modules target different but complementary aspects of the detection problem: computational efficiency, spatial-detail preservation, and multi-frequency feature representation. Their integration therefore forms a unified lightweight architecture designed specifically to balance detection capability and computational cost in UAV-based insulator inspection.

This integrated design is intended to preserve the feature information required for insulator detection while reducing the number of parameters and computational complexity relative to the baseline YOLOv11n architecture. The resulting lightweight architecture is designed with potential deployment on portable, embedded, and edge-computing platforms in mind, where model size and computational requirements are important practical constraints. The integrated architecture of the proposed lightweight YOLOv11n model is illustrated in Figure 8.

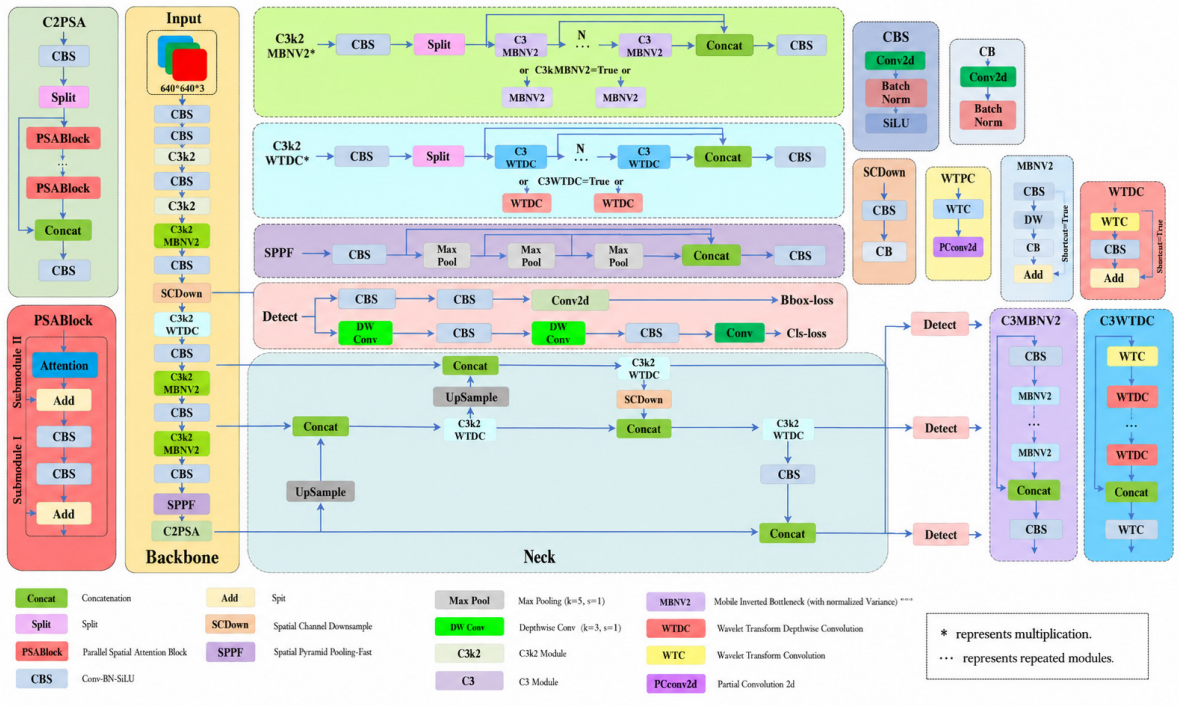


Figure 8. Integrated Architecture of the Proposed Lightweight YOLOv11 Model [21].

2.4. Evaluation Metrics

To evaluate the performance of the developed lightweight YOLOv11n model and to concretely reveal the impact of each structural improvement, a comprehensive performance analysis was conducted. The primary metrics used in this evaluation process include: Precision (P), Recall (R), Average Precision (AP), Mean Average Precision (mAP), and F1-score, as well as the total number of parameters (Params), floating-point operations (FLOPs), and inference time per image.

Among these metrics, Params and FLOPs quantitatively assess the computational complexity and hardware efficiency of the model, whereas P, R, AP, mAP, and F1-score are used to evaluate detection performance.

Within the scope of model accuracy analysis, correctly classified positive samples are defined as True Positives (TP), incorrect positive predictions as False Positives (FP), and missed positive samples as False Negatives (FN). In this study, TP represents cases where the model accurately predicts actual insulator regions; FP refers to instances where background regions are mistakenly detected as insulators; and FN denotes scenarios where the model fails to detect actual insulators.

In precision-critical applications such as insulator detection, the Precision (P) metric plays a vital role in minimizing false alarms and is calculated as follows:

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{1}$$

The recall (R) value represents the model's success in detecting positive objects. This metric is particularly important for reducing missed detections:

$$\mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2}$$

Average Precision (AP) represents the area under the precision–recall curve for a given class at a specified Intersection over Union (IoU) threshold. In this study, AP@0.5 denotes the AP calculated using an IoU threshold of 0.50. Mean Average Precision (mAP) is obtained by averaging the AP values across all target classes. Accordingly, mAP@0.5 represents the mean AP at an IoU threshold of 0.50, whereas mAP@0.5:0.95 represents the mean AP averaged over IoU thresholds ranging from 0.50 to 0.95 with a step size of 0.05. This metric represents the integration of accuracy and recall rates under different threshold values:

$$AP = \int_0^1 P(R)dR \tag{3}$$

If the evaluation is made on more than one class, the mAP (mean Average Precision) value is obtained by averaging the AP values of all classes:

$$mAP = \frac{1}{N}\sum_{i}^{N} AP_i \tag{4}$$

The F1 score is a critical metric that evaluates the balance between precision and recall in isolator detection models. It combines both metrics by balancing their values to provide a comprehensive performance evaluation. The F1 score is calculated as the harmonic mean of precision and recall, as defined in Equation (5):

$$F1\ score = \frac{2\mathrm{PR}}{\mathrm{P} + \mathrm{R}} \tag{5}$$

The mAP value obtained in this study demonstrates the model's ability to accurately detect isolators of varying sizes and locations and is directly related to overall performance. The developed model is optimized based on these metrics to deliver higher detection accuracy and lower computational overhead, particularly in complex scene structures. In this respect, the proposed architecture demonstrates balanced performance in terms of both accuracy and hardware efficiency, making it a powerful alternative for low-latency, real-time applications.

3. Results and Discussion

To maximize the performance of the proposed insulator detection model, a high-quality, custom-designed dataset was utilized. This dataset comprises a total of 5750 annotated images captured from various scenes containing insulators for the purpose of object detection. The images were collected from different geographical regions, under diverse environmental conditions (e.g., daylight, shaded areas, varying lighting intensities), and at various altitude levels. This diversity was intentionally incorporated to enhance the model's generalization capability.

To enable a controlled evaluation of model performance, the dataset was divided into training, validation, and test subsets using an 8:1:1 ratio. Specifically, 4600 images were allocated to the training set, while 575 images were assigned to the validation set and 575 images to the test set. The training subset was used for parameter optimization, the validation subset was used for model selection and hyperparameter monitoring, and the test subset was reserved exclusively for the final evaluation of the trained models.

To minimize the risk of data leakage caused by the high visual similarity of UAV images acquired from the same transmission-line structures, the dataset partitioning was performed at the acquisition-group level rather than solely at the individual-image level. Images originating from the same transmission tower or the same UAV acquisition sequence were assigned exclusively to one subset and were not distributed across the training, validation, and test sets. Consequently, visually similar observations from the same physical structure or acquisition sequence were prevented from appearing simultaneously in the training and evaluation subsets.

To increase the variability of the training data and improve robustness to changes in UAV viewing conditions, data augmentation was applied only to the training subset. The augmentation pipeline included random horizontal flipping, moderate geometric transformations, scaling, translation, and photometric variations such as brightness and contrast adjustment. These transformations were used to simulate variations in camera viewpoint, object scale, illumination, and positioning that may occur during UAV-based transmission-line inspection. No augmentation was applied to the validation or test subsets, thereby ensuring that performance metrics were calculated using unmodified evaluation images.

During the training phase, the model was trained based on the proposed lightweight YOLOv11 architecture, equipped with the optimized structural modules (C3k2MBNV2, SCDown, and C3k2WTDC). The training was conducted in a high-performance GPU environment, and the training hyperparameters (such as batch size, learning rate, and number of epochs) were meticulously fine-tuned based on validation performance. The experimental hardware and software setup, along with the key training parameters, are presented in detail in Table 1.

Figure 9 illustrates the changes in precision, recall, mAP, and F1-score throughout the training process. At the beginning of training, both precision and recall values gradually increase. Due to the high learning rate, mAP and F1-score initially rise rapidly but exhibit

noticeable fluctuations. After approximately the 50th epoch, these fluctuations begin to diminish, and the metrics start to improve in a more stable manner. Around the 150th epoch, the mAP and F1-score converge, indicating that the model has reached a stable performance level. The obtained results demonstrate that the model is capable of rapidly detecting insulators on high-voltage power lines and is suitable for real-time applications.

Table 1. Hardware, software, and training configurations used in the experiments.

Experimental Conditions		Key Parameters of Experimental Training	
Experimental Environment	**Details**	**Training Parameters**	**Details**
Programming language	Python 3.10	Epochs	200
Operating system	Ubuntu 22.04 TLS	Batch size	16
Deep learning framework	Pytorch 2.1.0	Image size (pixels)	640 × 640
CPU	Intel Core i9-13900KF	Optimization algorithm	AdamW
GPU	NVIDIA RTX 4080	Pre-training weights	None

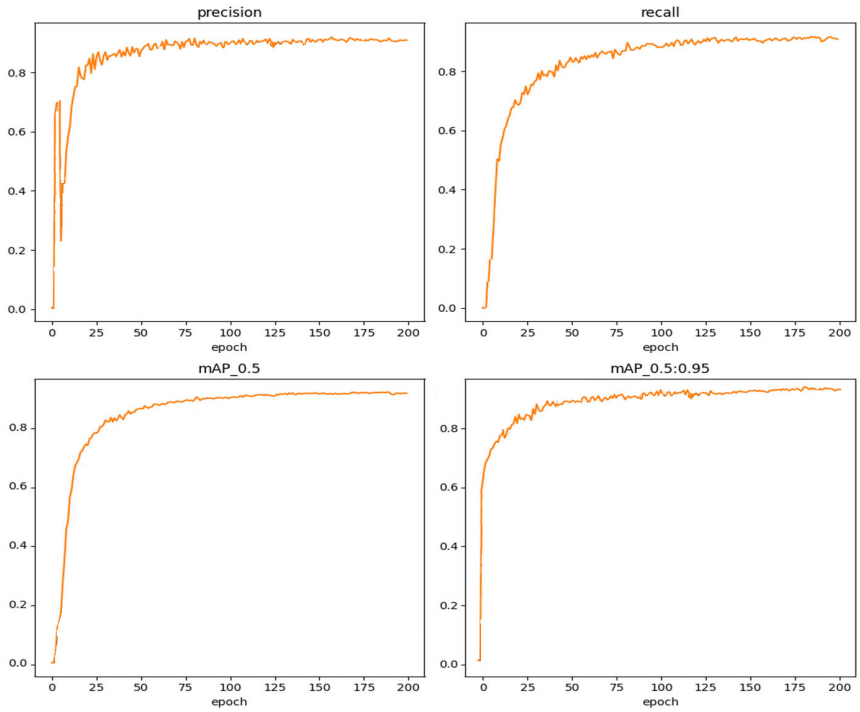


Figure 9. Training progress curve: precision of our model, recall of our model, mAP50 of our model, F1-score of our model.

3.1. Ablation Experiments

Figure 9 presents the model performance at each stage of the ablation study, based on the mAP curves of four different models. Upon examining the graph, it is observed that the proposed model exhibits a stable convergence trend after the 50th epoch and demonstrates a significant superiority in terms of mAP compared to the other models. This success can be attributed to the synergistic effect of the proposed blocks, which enhances the capacity for multi-scale feature fusion and effectively suppresses background noise in high-voltage transmission line environments.

These improvements enable accurate detection of insulators even in challenging scenes involving poor visibility, glare, shadowing, or other environmental difficulties. In particular, the C3k2WTDC block contributes to sharper and clearer detection of insulator edges, while the SCDown block helps preserve crucial structural information during resolution reduction. As a result of these modules, the model achieves higher accuracy in identifying

insulators on high-voltage power lines and enhances overall performance in detection tasks. This capability is especially critical in areas with complex infrastructure or in challenging scenarios where insulators overlap with background elements, as it significantly reduces the false positive rate during real-time monitoring and maintenance processes. Figure 10 shows the mAP50 comparison between different models.

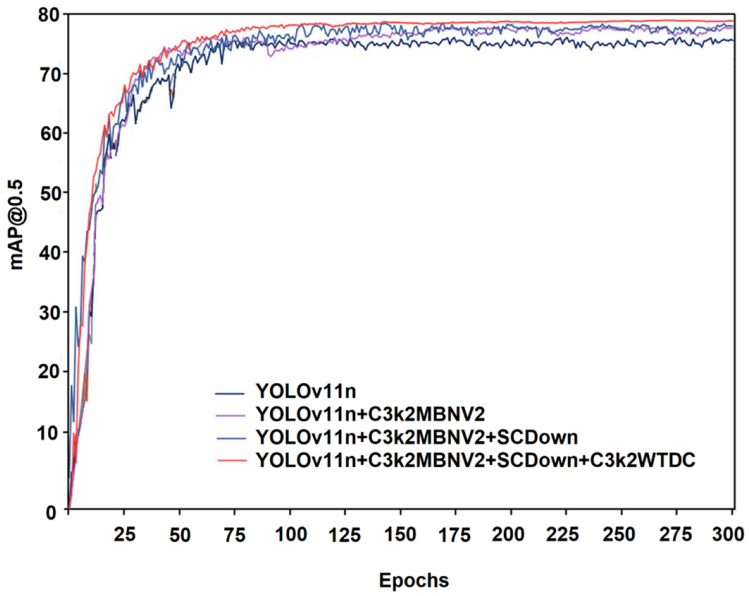


Figure 10. The comparison of mAP50 across different models is shown in the figure above.

Thanks to the synergistic effect of these modules, the model not only produced highly accurate results but also demonstrated consistent performance in challenging insulator detection tasks involving visual noise. The ability to correctly identify insulator structures even under complex environmental conditions such as dense vegetation and limited visibility is of great importance for real-time applications. Moreover, the experimental results confirm that the model's reliability in practical use has been enhanced by reducing the false positive rate.

Overall, each of the proposed structural blocks has improved different aspects of the model, resulting in a holistic performance enhancement. This has enabled successful insulator detection even in difficult scenes.

To systematically isolate the individual contributions and interaction effects of the proposed architectural components, a full factorial ablation study was conducted using C3k2MBNV2 (A), SCDown (B), and C3k2WTDC (C). In addition to the baseline YOLOv11n model, each module was evaluated individually, followed by all pairwise combinations and the complete A + B + C configuration. All configurations were evaluated under the same experimental protocol. Precision, recall, F1-score, mAP50, mAP50–95, inference time, parameter count, and GFLOPs were considered to jointly assess detection performance and computational efficiency. The quantitative results of the study are presented in detail in Table 2.

As shown in Table 2, the factorial ablation analysis reveals distinct contributions from the three architectural components. The baseline YOLOv11n model contains 6.20 M parameters and requires 20.5 GFLOPs, achieving a precision of 91.3%, a recall of 73.4%, an mAP50 of 71.2%, and an mAP50–95 of 65.6%. When C3k2MBNV2 (A) is introduced independently, the parameter counts decreases to 5.36 M and the computational complexity decreases to 17.6 GFLOPs, while precision increases slightly to 91.7%. This result indicates that the primary contribution of C3k2MBNV2 is the reduction in model complexity while maintaining detection performance.

Table 2. Factorial ablation analysis of the proposed architectural components.

Model	C3k2MBNV2	SCDown	C3K2WTDC	P (%)	R (%)	F1 (%)	mAP50 (%)	mAP50–95 (%)	Time (ms)	Params (M)	GFLOPs
YOLOv11n	-	-	-	91.3	73.4	81.4	71.2	65.6	2.7	6.20	20.5
A	√	-	-	91.7	73.7	81.9	71.7	65.9	2.9	5.36	17.6
B	-	√	-	92.2	74.5	82.1	72.3	66.3	2.9	5.75	16.7
C	-	-	√	92.0	74.2	82.1	72.5	66.6	3.1	5.15	17.3
A + B	√	√	-	92.6	74.5	82.5	73.0	66.8	3.1	4.26	14.3
A + C	√	-	√	92.9	74.5	82.5	73.3	67.0	3.2	4.04	14.9
B + C	-	√	√	93.1	75.1	83.1	73.5	67.1	3.2	4.20	14.2
Proposed	√	√	√	93.8	75.2	83.5	73.9	67.3	3.3	3.26	12.7

Note: "√" indicates that the corresponding module is included in the model configuration; "-" indicates that the module is not included.

The independent evaluation of SCDown (B) results in a precision of 92.2%, a recall of 74.5%, and an mAP50 of 72.3%, while reducing the computational complexity to 16.7 GFLOPs. Compared with the baseline, these results indicate that SCDown contributes not only to computational efficiency but also to improved feature preservation during down-sampling. In contrast, the independent C3k2WTDC configuration (C) achieves an mAP50 of 72.5% and an mAP50–95 of 66.6%. The improvement in mAP50–95 relative to the baseline suggests that the wavelet-based module contributes to more effective representation of fine and multi-frequency information across stricter IoU thresholds.

The pairwise experiments further demonstrate the complementary behaviour of the proposed components. The A + B configuration reduces the model to 4.26 M parameters and 14.3 GFLOPs while achieving an mAP50 of 73.0%. Combining C3k2MBNV2 with C3k2WTDC (A + C) further increases mAP50 and mAP50–95 to 73.3% and 67.0%, respectively. Among the pairwise configurations, B + C provides the strongest overall detection performance, achieving 93.1% precision, 75.1% recall, 73.5% mAP50, and 67.1% mAP50–95. These results indicate that SCDown and C3k2WTDC provide complementary spatial- and frequency-domain feature processing capabilities.

The complete A + B + C configuration provides the most favourable overall accuracy–complexity trade-off. The proposed model achieves a precision of 93.8%, a recall of 75.2%, an F1-score of 83.5%, an mAP50 of 73.9%, and an mAP50–95 of 67.3%, while requiring only 3.26 M parameters and 12.7 GFLOPs. Relative to the baseline YOLOv11n, this corresponds to approximately 47.4% fewer parameters and 38.0% fewer FLOPs, together with improvements of 2.5 percentage points in precision, 1.8 percentage points in recall, 2.7 percentage points in mAP50, and 1.7 percentage points in mAP50–95. Although inference latency increases moderately from 2.7 ms to 3.3 ms, the substantial reductions in parameter count and computational complexity, together with the improved detection metrics, demonstrate a favourable overall trade-off. Importantly, the factorial results show that the three components perform complementary functions rather than providing redundant modifications, with C3k2MBNV2 primarily contributing to model compression, SCDown supporting spatial-information preservation, and C3k2WTDC enhancing multi-frequency feature representation.

Although the proposed architecture substantially reduces the parameter count and computational complexity, the inference time increases from 2.7 ms for the baseline YOLOv11n to 3.3 ms for the complete model, corresponding to an absolute increase of 0.6 ms (approximately 22.2%). This result indicates that reductions in parameter count and FLOPs do not necessarily translate directly into lower latency. The additional wavelet-based feature decomposition and reconstruction operations in C3k2WTDC, together with the modified down-sampling operations, introduce execution overhead that is not fully reflected by FLOPs alone. Nevertheless, the resulting latency of 3.3 ms per image remains low in the GPU-based experimental environment used in this study.

The results indicate that the proposed architecture not only enhances accuracy performance but also significantly optimizes resource consumption, thereby improving its applicability to systems with limited hardware capacity. The final model's memory footprint was reduced to nearly half that of the original YOLOv11n architecture. This provides a substantial advantage for deployment on low-power edge devices, particularly in field conditions where insulator detection is required under harsh environmental constraints.

3.2. Performance Evaluation

To objectively demonstrate the improvements in detection accuracy of the model, the developed framework was evaluated on a custom test set and the results were compared with existing baseline methods. The achieved performances are detailed in Table 3.

Table 3. Performance comparison of lightweight YOLO models evaluated on the same test dataset.

Model	Precision%	Recall%	F1-Score%	mAP50%	mAP50–95%	Time (ms)	Params (P)	GFLOPs
YOLOv5s	75.3	60.2	66.7	63.7	61.2	2.3	2.7	7.7
YOLOv7	84.3	67.2	74.7	65.8	63.1	2.4	2.6	8.3
YOLOv8n	81.5	63.9	71.4	65.1	62.0	2.5	3.1	8.7
YOLOv9t	85.4	70.1	76.9	67.2	65.0	2.9	2.0	7.6
YOLOv10n	89.6	72.5	79.7	70.5	65.2	2.7	2.4	6.6
YOLOv11n	91.3	73.4	81.4	71.2	65.6	2.7	6.2	20.5
Proposed YOLOv11n	93.8	75.2	83.5	73.9	67.3	3.3	3.26	12.7

As shown in Table 3, the proposed lightweight YOLOv11n architecture achieves a favourable balance between detection performance and computational complexity. Compared with the baseline YOLOv11n, the proposed model reduces the number of parameters from 6.20 M to 3.26 M and the computational complexity from 20.5 to 12.7 GFLOPs, corresponding to reductions of approximately 47.4% and 38.0%, respectively. At the same time, precision increases from 91.3% to 93.8%, recall from 73.4% to 75.2%, mAP50 from 71.2% to 73.9%, and mAP50–95 from 65.6% to 67.3%. The comparison with other lightweight YOLO variants further provides a controlled assessment of the trade-off between model complexity and detection performance under the same experimental conditions.

In experiments conducted under challenging environmental conditions, such as real-time insulator detection scenarios, the developed model demonstrated more balanced and comprehensive detections. Particularly in complex scenes with cluttered backgrounds, the model exhibited high-precision detection capability. To evaluate the model's effectiveness, comparative experiments were conducted against widely used real-time architectures including YOLOv5s, YOLOv7, YOLOv8n, YOLOv9t, YOLOv10n, and YOLOv11n. In addition to the controlled comparison in Table 3, selected previously reported lightweight and insulator-related detection methods are summarized in Table 4 only to provide broader contextual information regarding model complexity and reported detection performance.

Table 4 provides a contextual comparison between the proposed architecture and previously reported detection methods. Because the literature results were obtained using different datasets, tasks, input resolutions, training protocols, and hardware platforms, the reported numerical values are not interpreted as direct evidence of superiority. The model achieved a precision of 93.8% and a recall of 75.2%, presenting a well-balanced performance profile. The architectural design of the model is optimized not only for accuracy but also for computational efficiency. With a total of 3.26 million parameters and 12.7B FLOPs, the model achieves inference with only 3.3 ms latency per image. This low latency enables the deployment of the model on edge devices and platforms with limited hardware resources.

Table 4. Contextual comparison of the proposed model with previously reported insulator detection methods.

Authors	Model	Dataset/Application	mAP50 (%)	Params (M)	GFLOPs
Huang et al. [4]	TLI-YOLOv5	UAV transmission-line inspection	56.6	-	4.2
Qin et al. [11]	Improved YOLOv8	CPLID	98.6	3.05	-
Li et al. [25]	Lite YOLO-ID	IDID-Plus	65.1	3.71	8.7
Deng et al. [26]	RML-YOLO	UID/UAV insulator defect detection	88.3	2.51	-
Lu et al. [27]	IDD-YOLO	UAV insulator defects	66.2	2.85	5.1
Zhang et al. [28]	Improved YOLOv5	Insulator defects	99.4	3.81	9.9
This study	YOLOv11n (baseline)	Custom UAV dataset	71.2	6.20	20.5
This study	Proposed YOLOv11n	Custom UAV dataset	73.9	3.26	12.7

Note: The YOLOv11n baseline and the proposed model labelled as "This study" were evaluated under the same dataset split, input resolution, training protocol, and hardware environment and therefore provide a directly controlled comparison. Other literature results are included only for contextual reference because they were obtained under different experimental conditions.

Considering both its high mAP and low resource consumption, it can be concluded that the developed system offers an effective and feasible solution for real-time insulator detection applications. In Table 4, a comparative analysis of multiple isolator detection algorithms is shown.

Although several studies report higher absolute mAP50 values, these results were obtained using different datasets and experimental protocols and therefore cannot be directly compared with the present results. The primary contribution of the proposed architecture is demonstrated by the controlled experiments in Table 3, where it improves detection performance over the baseline YOLOv11n while reducing the parameter count by approximately 47.4% and GFLOPs by 38.0%.

Within the controlled comparison performed in this study, the proposed architecture improves precision from 91.3% to 93.8% and recall from 73.4% to 75.2% relative to the baseline YOLOv11n. At the same time, the parameter counts decreases from 6.20 M to 3.26 M and the computational complexity decreases from 20.5 to 12.7 GFLOPs.

The proposed model was also compared contextually with previously reported insulator detection approaches to situate its computational and detection characteristics within the existing literature. As summarized in Table 4, the proposed architecture maintains effective detection performance while operating with a relatively lightweight computational structure. However, the reported results were obtained under different datasets, data distributions, input resolutions, training protocols, and hardware configurations. Therefore, these numerical values are presented only for contextual reference and should not be interpreted as evidence of direct superiority among the compared methods. Instead, the comparison is intended to illustrate the range of accuracy–efficiency trade-offs reported in recent insulator detection studies and to provide a broader context for the design characteristics of the proposed model.

Beyond performance metrics, one of the most significant distinguishing features of the proposed model is its ability to achieve high detection accuracy with a remarkably low number of parameters and computational cost. For example, the proposed architecture requires substantially fewer parameters and FLOPs than several larger models reported in the literature. Nevertheless, these complexity and detection metrics should be interpreted in the context of the different experimental conditions used in the respective studies. These observations provide contextual evidence that the proposed architecture maintains

relatively low model complexity; however, direct efficiency comparisons remain limited by differences in experimental settings across studies.

From a real-time performance perspective, the model has been tested on an NVIDIA RTX 4080 GPU platform, where it achieved high FPS rates, confirming that its inference time is fast enough for real-world applications. The SCDown module's efficient down-sampling structure and the wavelet-based transformation module's ability to mitigate information loss contribute to the model's robust performance even in complex or low-quality visual environments.

Overall, the literature comparison should be interpreted as contextual rather than directly comparative. The controlled experiments reported in Table 3 provide the primary evidence for assessing the relative performance of the proposed YOLOv11n architecture, while Table 4 positions the model within the broader range of previously reported approaches.

3.2.1. Small-Scale Insulator Detection Performance

The detection of small-sized or partially occluded insulators located on high-voltage power lines poses a significant challenge, particularly under complex environmental conditions. In this study, the performance of the developed lightweight YOLOv11n-based model was tested using a custom dataset specifically designed to evaluate its capability in detecting small-scale insulator components. The test images used in the evaluation process were not included in the training dataset.

As illustrated in Figure 11, the model trained with the proposed dataset demonstrated high accuracy in detecting small and detailed insulator structures. This result indicates that the proposed dataset provides a better representation of complex scenarios and effectively reduces the false negative rate.

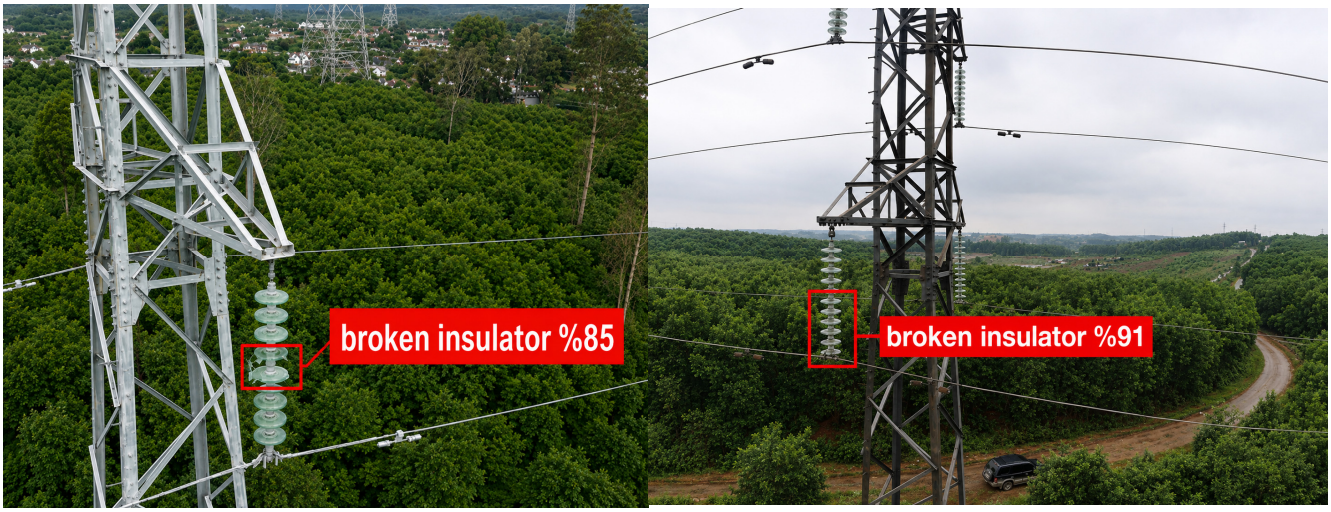


Figure 11. Detection results of small and partially covered insulators.

3.2.2. Distinguishing Cloud-like Structures

During isolator inspection, particularly in UAV imagery, the confusion of visual elements like clouds or smoke resulting from environmental influences with dirty isolator surfaces can lead to false-positive detections. To assess the impact of this issue, the developed model was trained on the proposed dataset, and a comparative analysis was conducted. The images selected for testing consisted of samples containing both smoke and clouds, which were not included in the training data.

The images presented in Figure 12 demonstrate that the model trained with the proposed dataset can accurately detect isolators and does not confuse similar formations

like clouds or smoke. This finding demonstrates that the model is more resilient to visual similarities and delivers more reliable results in real-world situations.

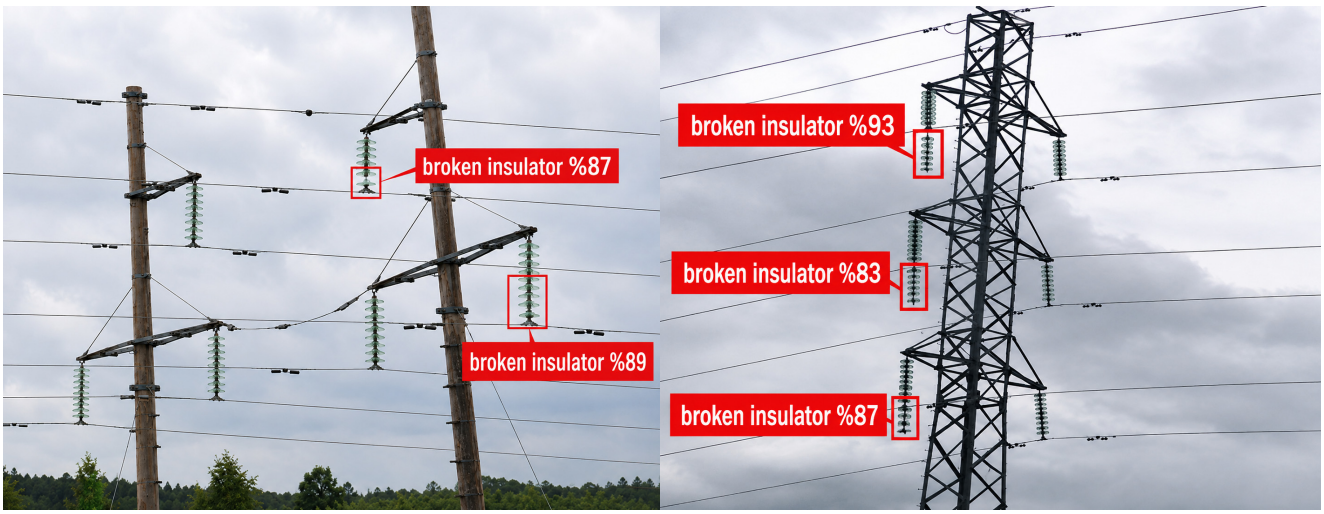


Figure 12. Model resilience analysis against cloud effects.

3.2.3. Consistency of the Model in Different Environmental Conditions

For real-time isolator monitoring systems to be practical, they must be able to operate stably in various lighting and weather conditions. The model's performance was evaluated in various scenarios, including daytime, night-time, cloudy, and hazy weather. The images used during testing were not included in the training process.

The examples in Figure 13 demonstrate that the model successfully detected isolator structures and potential defects under all conditions. It was determined that the model performed with high accuracy, particularly in low-light night-time images and in hazy environments with intense visual noise. This demonstrates the model's robust ability to adapt to varying scene conditions.



Figure 13. Detection consistency of the model in different environmental conditions.

The model's ability to operate with low false positive and negative rates, even in such challenging environmental conditions, is crucial for early detection and intervention. Furthermore, this capability increases the model's applicability not only in the field but also across various sectors. It demonstrates that it offers a robust technical infrastructure, particularly for long-term monitoring and use with endpoint devices.

4. Conclusions

This study investigated whether a lightweight YOLOv11n architecture can address the competing requirements of detection performance and computational efficiency in UAV-based high-voltage insulator inspection. The proposed architecture integrates C3k2MBNV2, SCDown, and C3k2WTDC to address three complementary aspects of the detection problem: model complexity, spatial-information preservation during down-sampling, and multi-frequency feature representation. The experimental results indicate that these modifications can substantially reduce the computational requirements of the baseline architecture while maintaining competitive detection performance.

Beyond the numerical improvements, the results suggest that lightweight model design for UAV-based insulator inspection should not rely solely on parameter reduction. Preserving spatial information is particularly important when insulators occupy a small portion of aerial images, while frequency-domain representations can provide complementary information for fine local structures. The factorial ablation analysis further enables the individual and combined effects of the three architectural components to be distinguished, providing a clearer interpretation of the trade-off between model complexity and detection capability.

The findings also have practical implications for automated transmission-line inspection. A detector with reduced parameter and computational requirements can lower the computational burden associated with processing large volumes of UAV imagery and may facilitate future deployment on resource-constrained platforms. However, the present results should be interpreted within the scope of the experimental setting. The model was evaluated using the dataset and hardware environment considered in this study, and direct on-device validation on embedded UAV hardware was not performed. In addition, broader validation across independent datasets, different transmission-line infrastructures, camera systems, and geographic regions is required to establish the generalizability of the proposed architecture.

Future work will therefore focus on validating the model on independent datasets and resource-constrained edge hardware, extending the framework from image-based detection to video-based tracking, and integrating the detection system with autonomous UAV inspection routes. Further investigation of feature responses and frequency-domain behavior may also provide deeper insight into the mechanisms through which the proposed modules contribute to small-object insulator detection.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

YOLO	You Only Look One
UAV	Unmanned Aerial Vehicle
CNN	Convolutional Neural Network
FLOPs	Floating Point Operations
mAP	Mean Average Precision
C3k2MBNV2	Cross-Stage Partial Narrow Bottleneck with MobileNetV2
SCDown	Spatial-Channel Decoupled Downsampling
C3k2WTDC	Cross-Stage Partial Narrow Bottleneck with Wavelet Transform Depthwise Convolution

DWConv	Depthwise Convolution
FPN	Feature Pyramid Network
PANet	Path Aggregation Network

References

1. Wei, L.; Jin, J.; Deng, K.; Liu, H. Insulator Defect Detection in Transmission Line Based on an Improved Lightweight YOLOv5s Algorithm. *Electr. Power Syst. Res.* **2024**, *233*, 110464. [CrossRef]
2. Chen, R.; Lui, Z.; Ou, W.; Zhang, K. Small target detection algorithm based on improved YOLOv5. *Electronics* **2024**, *13*, 4158. [CrossRef]
3. Liu, C.; Wu, Y.; Liu, J.; Sun, Z.; Xu, H. Insulator Faults Detection in Aerial Images from High-Voltage Transmission Lines Based on Deep Learning Model. *Appl. Sci.* **2021**, *11*, 4647. [CrossRef]
4. Huang, H.; Lan, G.; Wei, J.; Zhong, Z.; Xu, Z.; Li, D.; Zou, F. TLI-YOLOv5: A Lightweight Object Detection Framework for Transmission Line Inspection by Unmanned Aerial Vehicle. *Electronics* **2023**, *12*, 3340. [CrossRef]
5. Sadykova, D.; Pernebayeva, D.; Bagheri, M.; James, A. IN-YOLO: Real-Time Detection of Outdoor High Voltage Insulators Using UAV Imaging. *IEEE Trans. Power Deliv.* **2020**, *35*, 1599–1601. [CrossRef]
6. Yin, Y.; Liu, X.; Wu, K.; Xu, R.; Zheng, J.; Mei, F. Lightweight Power Line Defect Detection Based on Improved YOLOv8n. *Sensors* **2026**, *26*, 2112. [CrossRef] [PubMed]
7. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
8. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767. [CrossRef]
9. Zhao, Z.-Q.; Zheng, P.; Xu, S.-T.; Wu, X. Object Detection with Deep Learning: A Review. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 3212–3232. [CrossRef] [PubMed]
10. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934. [CrossRef]
11. Qin, Y.; Zeng, Y.; Wang, X. Insulator Defect Detection in Complex Environments Based on Improved YOLOv8. *Entropy* **2025**, *27*, 633. [CrossRef] [PubMed]
12. You, X.; Zhao, X. A Insulator Defect Detection Network Based on Improved YOLOv7 for UAV Aerial Images. *Measurement* **2025**, *253*, 117410. [CrossRef]
13. Jiang, T.; Hou, X.; Wang, M. Insulator Defect Detection Based on the CDDCR–YOLOv8 Algorithm. *Int. J. Comput. Intell. Syst.* **2024**, *17*, 245. [CrossRef]
14. Wang, Z.; Yuan, G.; Zhou, H.; Ma, Y.; Ma, Y.; Chen, D. Improved YOLOv7 Model for Insulator Defect Detection. *Electron. Res. Arch.* **2024**, *32*, 2880–2896. [CrossRef]
15. Han, W.; Cai, Z.; Li, X.; Ding, A.; Zou, Y.; Wang, T. LMD-YOLO: A Lightweight Algorithm for Multi-Defect Detection of Power Distribution Network Insulators Based on an Improved YOLOv8. *PLoS ONE* **2025**, *20*, e0314225. [CrossRef] [PubMed]
16. Zhao, J.; Miao, S.; Kang, R.; Cao, L.; Zhang, L.; Ren, Y. Insulator Defect Detection Algorithm Based on Improved YOLOv11n. *Sensors* **2025**, *25*, 1327. [CrossRef] [PubMed]
17. Dian, S.; Zhong, X.; Zhong, Y. Faster R-Transformer: An Efficient Method for Insulator Detection in Complex Aerial Environments. *Measurement* **2022**, *199*, 111238. [CrossRef]
18. Hien, A.T.; Tran, A.D.; Viet, D.C.; Thuy, Q.D.T.; Huu, Q.N. Insulator Defect Detection Based on Feature Pyramid Network and Diffusion Model. *Signal Image Video Process.* **2025**, *19*, 372. [CrossRef]
19. Ultralytics. YOLOv11 GitHub Repository. 2024. Available online: https://github.com/ultralytics (accessed on 29 June 2025).
20. Jia, N.; Wei, Z.; Li, B. Attention-Enhanced Lightweight One-Stage Detection Algorithm for Small Objects. *Electronics* **2023**, *12*, 1607. [CrossRef]
21. Tao, Y.; Li, B.; Li, P.; Qian, J.; Qi, L. Improved Lightweight YOLOv11 Algorithm for Real-Time Forest Fire Detection. *Electronics* **2025**, *14*, 1508. [CrossRef]
22. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520. [CrossRef]
23. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J.; Ding, G. YOLOv10: Real-Time End-to-End Object Detection. *arXiv* **2024**, arXiv:2405.14458. [CrossRef]
24. Zhang, X.; Wang, Z.; Zhou, H.; Ghosh, S.; Gnanapragasam, D.; Jampani, V.; Su, H.; Guibas, L. CONDENSE: Consistent 2D/3D Pre-training for Dense and Sparse Features from Multi-View Images. In *Computer Vision—ECCV 2024*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2024; pp. 19–38. [CrossRef]
25. 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. Instrum. Meas.* **2024**, *73*, 5023812. [CrossRef]

26. Deng, Z.; Li, X.; Yang, R. RML-YOLO: An Insulator Defect Detection Method for UAV Aerial Images. *Appl. Sci.* **2025**, *15*, 6117. [CrossRef]
27. Lu, Y.; Li, D.; Li, D.; Li, X.; Gao, Q.; Yu, X. A Lightweight Insulator Defect Detection Model Based on Drone Images. *Drones* **2024**, *8*, 431. [CrossRef]
28. Zhang, T.; Zhang, Y.; Xin, M.; Liao, J.; Xie, Q. A Light-Weight Network for Small Insulator and Defect Detection Using UAV Imaging Based on Improved YOLOv5. *Sensors* **2023**, *23*, 5249. [CrossRef] [PubMed]

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