Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 1, February 2026, Page. 60-80
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4679

Insulator Defect Detection Based On Image Processing Using A Modified YOLOv8n Model

Muchamad Arfim Muzaki[1], Subiyanto*[2], Andika Anantyo[3]

[1,2,3]Department of Electrical Engineering, Faculty of Engineering, Universitas Negeri Semarang, Indonesia

Email: [2]subiyanto@mail.unnes.ac.id

Received : May 3, 2025; Revised : Jun 24, 2025; Accepted : Jun 25, 2025; Published : Feb 15, 2026

Abstract

Insulators are critical components in power transmission and distribution systems, where any defects can lead to severe operational failures and power outages. To enhance inspection efficiency, unmanned aerial vehicles (UAVs) are increasingly used for aerial monitoring. However, the quality of images captured by drones is often compromised due to hardware limitations, motion blur, and complex environmental backgrounds, which significantly reduces the performance of deep learning-based defect detection methods. This study proposes an improved insulator defect detection model based on the YOLOv8n architecture, optimized for accuracy and efficiency in low-quality image scenarios and suitable for deployment in resource-constrained environments. The model introduces two major modifications. First, a Slim-Neck module employing Ghost-Shuffle Convolution (GSConv) replaces standard convolutions to substantially reduce computational cost while preserving rich feature representations. Second, an Efficient Multi-Scale Attention (EMA) module is integrated into the neck to enhance multi-scale feature fusion by maintaining per-channel information without dimensionality reduction, improving the model's ability to extract discriminative features. Experimental results demonstrate that the proposed model achieves a precision of 92.0%, recall of 88.6%, mAP@0.5 of 92.1%, and an inference speed of 161.29 FPS. Furthermore, it reduces parameter count by 10.8% and computational load by 8.6% compared to the baseline, validating its suitability for real-time UAV-based inspections. The model also outperforms existing methods in detecting insulator defects, particularly in challenging conditions involving blur and complex backgrounds.

Keywords : *Insulator Defect Detection, YOLOv8, Slim-Neck, Efficient Multi-Scale Attention (EMA).*

1. INTRODUCTION

As the power grid continues to expand, transmission line installation areas are also growing [1]. To ensure long-term operational stability, regular inspections are required to detect potential defects in various components [2]. Insulators, as one of the most critical components, play a vital role in supporting conductors and maintaining electrical insulation throughout the transmission process [3], [4]. Since they are typically installed outdoors, insulators are frequently exposed to extreme environmental conditions. This exposure makes their surfaces prone to contamination and susceptible to damage such as corrosion, cracking, and breakage [5]. Such deterioration can degrade insulation performance, potentially leading to system faults and reducing the overall reliability and efficiency of power system operations [6]. Moreover, insulators are among the components with the highest failure rates in transmission systems, and their malfunction can result in widespread system disruptions and significant economic losses [7], [8]. Therefore, early detection and timely replacement of defective insulators are crucial measures to maintain the reliability of power systems [9], [10].

Currently, defect detection in power system insulators still heavily relies on conventional methods such as manual visual inspections and handheld device usage [11]. These traditional methods are labor-

Vol. 7, No. 1, February 2026, Page. 60-80
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4679

intensive and prone to inconsistency due to human subjectivity. Moreover, the remote and complex environments of high-voltage transmission lines make manual inspections inefficient and impractical. [12]. With the advancement of aerial survey technologies, drone-based inspection of electrical infrastructure has become increasingly utilized [13]. Although this contactless detection method helps overcome the limitations of manual inspection, it still faces challenges such as varying viewpoints in aerial imagery, complex background interference, and uneven signal intensities [14].

Since manually reviewing drone imagery is no longer practical, many researchers have begun to adopt deep learning based detection algorithms for identifying defects in insulators [15]. In recent years, deep learning based insulator defect detection has become a major focus in research due to its superior capability in accurately detecting faults [16]. Deep learning methods have demonstrated remarkable performance in complex tasks such as image classification and component defect detection [17]. The primary advantages of deep learning over traditional methods include high accuracy, fast detection speed, automatic extraction of deep features, and resilience against environmental disturbances [18].

Deep-learning object detectors are generally classified into two categories. Two-stage, region-based methods such as R-CNN, Faster R-CNN, FPN, and Mask R-CNN offer high accuracy but are complex and relatively slow [19]. One-stage, regression-based methods such as YOLO, SSD, EfficientDet, and RetinaNet provide faster detection with simpler architectures, though at a slight cost in accuracy [20]. YOLO is widely adopted for insulator-fault detection due to its balance of speed and precision [21]. However, UAV-based inspections still face challenges such as cluttered backgrounds, diverse insulator shapes, small defect sizes, and poor image quality from long-range capture [16], [18]. To meet these challenges, recent studies have embedded lightweight convolutions and attention modules into YOLO architectures. For instance, the integration of CBAM into YOLOv5 enhances small-target selectivity [22]. In another approach, GhostConv combined with Coordinate Attention and an EVCBlock is used in YOLOv5s to enrich feature representation [23]. YOLOv7 has been modified by applying depthwise GhostConv along with SE attention, which helps reduce the number of parameters without compromising accuracy [24]. Similarly, a variant of YOLOv8n incorporates Triplet Attention, SCConv, a SC-Detect head, and a slim GSConv neck, leading to improved detection efficiency and overall performance [25].

Recent advancements in object detection have focused on combining lightweight convolutional operations with attention mechanisms to improve both accuracy and computational efficiency an essential requirement for edge deployment on UAV platforms. MI-YOLO [26] enhances YOLOv5s by incorporating attention modules and dilated convolutions, while similar improvements using CBAM [27], and EMA-integrated FasterNet [28] have demonstrated better recall for detecting small defects. Building on these trends, newer architectures such as LMD-YOLO [29], IALF-YOLO [30], and GSM-YOLO [31] further advance performance by integrating modules like SCConv, SimAM, and MAP-CA attention, achieving high detection speed and accuracy. However, despite these gains, many models still face difficulties under extreme lighting conditions, in cluttered backgrounds, or when detecting very small defects. In addition, the increasing complexity of these architectures can hinder real-time performance on resource-limited UAVs, while dependence on narrow or imbalanced datasets continues to limit their generalization to diverse real-world environments.

To overcome these limitations particularly in detecting small or subtle defects within low-resolution and noisy aerial images this study proposes an improved insulator defect detection framework based on the YOLOv8n architecture. The proposed model is optimized to enhance detection accuracy under challenging visual conditions while maintaining computational efficiency, making it well-suited for deployment in real-time and resource-constrained environments. The primary contributions of this work include: (1) the integration of an Efficient Multi-Scale Attention (EMA) module [32] enhances multi-scale spatial and channel-aware feature fusion, significantly improving the detection of insulator

Vol. 7, No. 1, February 2026, Page. 60-80
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4679

defects in complex industrial environments. (2) Reconstructing the neck structure with a GSConv-based Slim-Neck [33] optimizes the model's parameters and computational load while preserving detection accuracy, enabling efficient real-time deployment on resource-constrained edge devices. Through these innovations, the proposed model aims to improve real-time defect detection in drone inspection systems, facilitating more efficient monitoring of critical infrastructure and supporting sustainable power grid maintenance.

2. METHOD

To develop an efficient and accurate insulator defect detection system, a structured methodology was implemented, consisting of several key stages ranging from dataset preparation to model evaluation. Each stage was designed to ensure the successful development, training, and deployment of a YOLO-based model optimized for real-time detection in challenging environments. The research stages are illustrated in Figure 1.

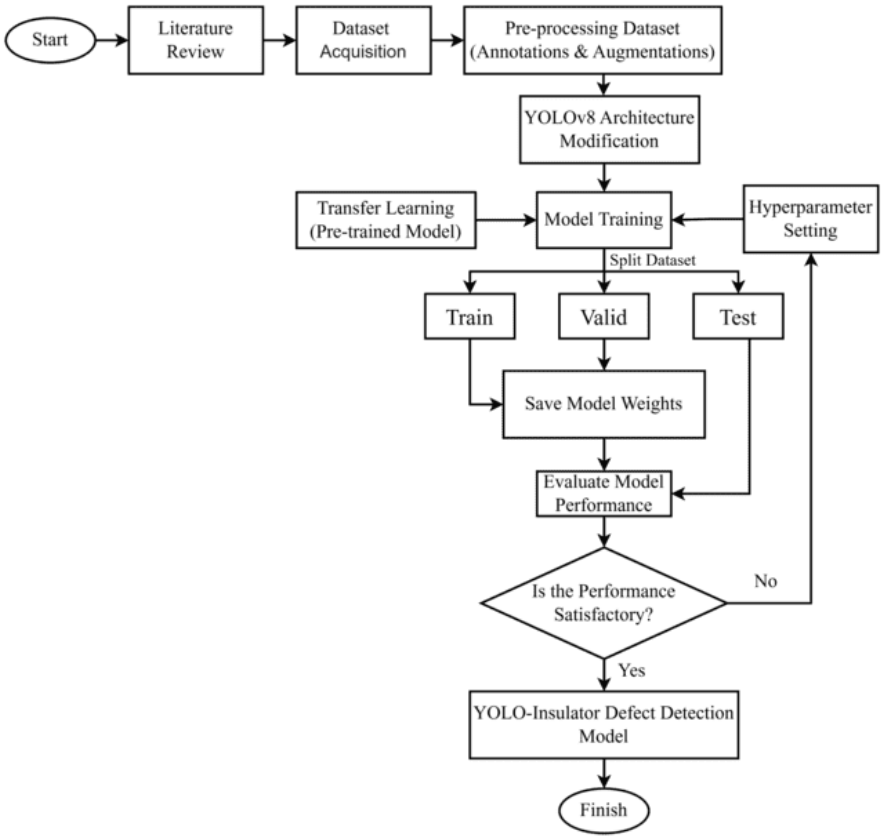

Figure 1. Research Workflow for the Development of the YOLO-Based Insulator Defect Detection System

The research began with a literature review, aimed at understanding the state-of-the-art methods in insulator defect detection, including various YOLO variants and improvements using lightweight convolutional and attention modules. Next, dataset acquisition was performed by collecting and curating image data of insulators, including both defective and non-defective conditions. The dataset was then processed in the pre-processing stage, where image annotations and data augmentations were applied to improve model generalization and balance the training set. The following phase involved YOLOv8 architecture modification. In this step, the standard YOLOv8n structure was modified by integrating two custom modules: (1) a Slim-Neck using GSConv for reduced model complexity and (2) an Efficient Multi-Scale Attention (EMA) module to enhance feature representation at different spatial and channel

Vol. 7, No. 1, February 2026, Page. 60-80
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4679

levels. Transfer learning was then applied using pre-trained weights as the model's initial parameters, providing a strong starting point and reducing training time. Simultaneously, hyperparameter setting was conducted to define key parameters such as batch size, learning rate, and training epochs. The model was then subjected to the training phase, where the data was divided into train, validation, and test sets. During this phase, the model's parameters were optimized, and performance was monitored using metrics such as precision, recall, and mean average precision (mAP). Upon completion of training, the best-performing model weights were saved, followed by a comprehensive evaluation phase to assess detection accuracy and inference speed (FPS) using both quantitative metrics and qualitative visual comparisons. If the performance met the predefined benchmarks, the model was finalized as the YOLO-based Insulator Defect Detection Model, ready for deployment in automated inspection systems, especially on resource-constrained platforms such as drones or edge devices.

2.1. Proposed Framework for Insulator Defect Detection

To ensure accurate and efficient defect detection on insulator surfaces, this study proposes an improved object detection framework based on a modified YOLOv8n architecture. The proposed framework integrates lightweight convolutional modules and attention mechanisms to enhance both performance and deployability, especially in edge computing environments. The architecture leverages transfer learning and multi-scale feature fusion to detect fine-grained defects under various visual conditions.

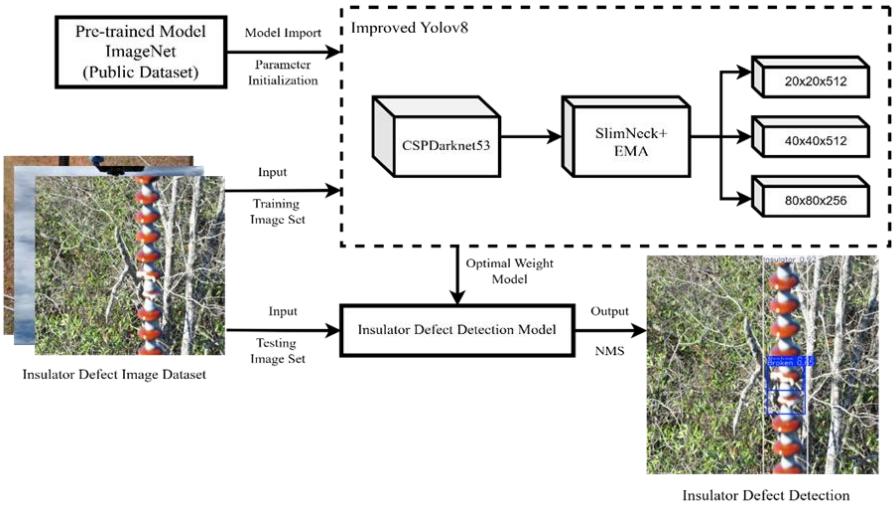

Figure 2. The proposed framework for insulator defect detection using a modified YOLOv8n model

As illustrated in Figure 2, the framework begins by inputting insulator images, which are processed through a feature extraction backbone based on CSPDarknet53. The model uses pre-trained weights from the ImageNet public dataset to initialize the backbone, enabling better convergence and generalization. The extracted feature maps are then passed through a SlimNeck module, which replaces the default YOLO neck with a more efficient version using GSConv to reduce model complexity while preserving spatial information. To further improve detection robustness, the Efficient Multi-Scale Attention (EMA) module is embedded within the neck. EMA consists of parallel spatial and channel attention paths that highlight relevant features and suppress irrelevant background noise. This combination helps the model remain sensitive to small-scale defects in complex visual environments. The final detection head processes the multi-resolution feature maps (20×20×512, 40×40×512, 80×80×256) to localize and classify insulator defects. The output is a detection result that accurately identifies defective regions with bounding boxes and confidence scores.

Vol. 7, No. 1, February 2026, Page. 60-80
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4679

2.2. Architectural Design and Core Components of YOLOv8

YOLOv8 represents a state-of-the-art evolution in the YOLO family of single-stage object detection frameworks, renowned for its real-time performance and balance between speed and accuracy. Building upon its predecessors, YOLOv8 employs a modular architecture comprising three core components: a backbone network for hierarchical feature extraction, a neck network for multi-scale feature fusion, and a detection head for precise localization and classification of objects. As shown in Figure 3, the backbone leverages the CSPDarknet architecture to efficiently capture spatial and semantic features across different scales, while the neck integrates advanced feature fusion techniques, such as PANet (Path Aggregation Network), to aggregate contextual information from multiple layers. The detection head further refines these features to predict bounding boxes and class probabilities with minimal computational overhead. This streamlined design enables YOLOv8 to excel in scenarios requiring rapid and reliable detection, such as identifying insulator defects in power grid infrastructure, where both real-time processing and high accuracy are critical for operational safety and efficiency.

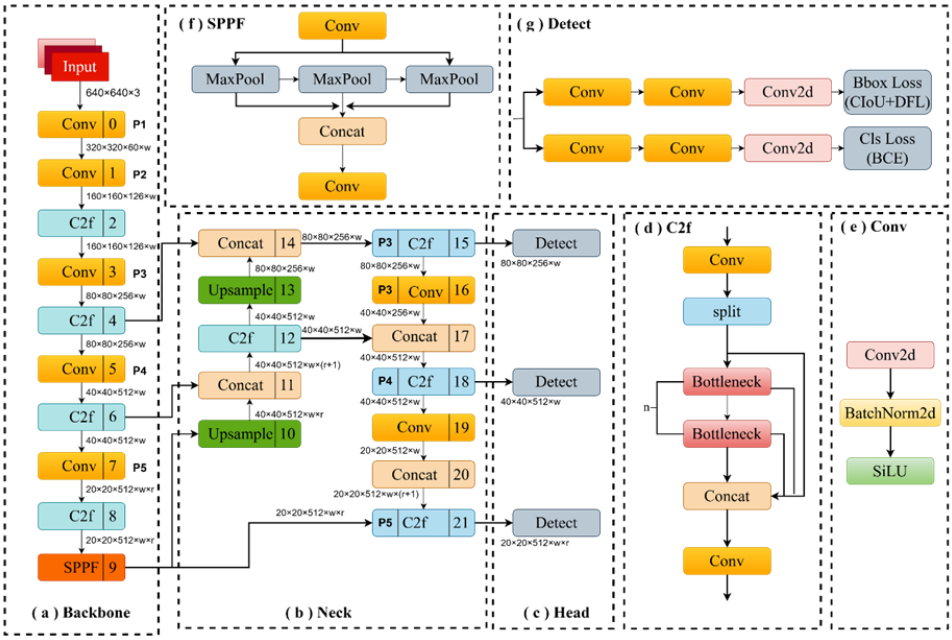

Figure 3. Structure of YOLOv8

Figure 3 presents the detailed architecture of the YOLOv8 model, illustrating its three core components: the Backbone, Neck, and Head. The Backbone is responsible for initial feature extraction and is composed of alternating convolutional (Conv) layers and C2f modules, which replace the traditional C3 structure to improve gradient flow and enhance the richness of extracted features. The spatial pyramid pooling fast (SPPF) module at the end of the Backbone further aggregates spatial information from different receptive fields. The Neck section employs a feature pyramid network (FPN)-like structure with upsampling and concatenation operations, enabling multi-scale feature fusion that enhances the model's ability to detect objects of varying sizes. The Head contains three detection branches (P3, P4, and P5), each targeting a specific scale of objects, and utilizes binary cross-entropy (BCE) for classification loss and a combination of Complete IoU (CIoU) and Distribution Focal Loss (DFL) for bounding box regression. YOLOv8 has several variants, ranging from YOLOv8n to YOLOv8x, which are designed to meet various object detection needs. In this work, YOLOv8n is chosen because of its lightweight and efficient model size, enabling fast object detection even with limited hardware resources.

Vol. 7, No. 1, February 2026, Page. 60-80
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4679

2.3. Architectural Design of the Modified YOLOv8n Model

While YOLO excels in rapid, real-time object detection, it faces challenges such as excessive computational expenses and reduced accuracy when identifying objects against intricate or busy backgrounds. To overcome these limitations, this research proposes an enhanced version of YOLOv8n, termed Modified-YOLOv8n, which implements two critical innovations. First, a Slim-Neck architecture utilizing GSConv is introduced to minimize computational demands without compromising the richness of feature representations. Second, an EMA module is integrated into the neck structure to boost the detection of insulator defects within visually cluttered scenes. As shown in Figure 4b, the improvements involve two structural changes: substituting the original C2f module with the efficient cross stage partial block (VoV-GSCSP) component and replacing standard convolutional layers with GSConv operations to streamline efficiency. Moreover, EMA modules are strategically implemented after the final P3, P4, and P5 VoV-GSCSP layers to sharpen feature refinement and amplify the model's focus on insulator defects, even in environments dominated by complex background interference. These adaptations collectively enhance both computational efficiency and detection precision in challenging scenarios.

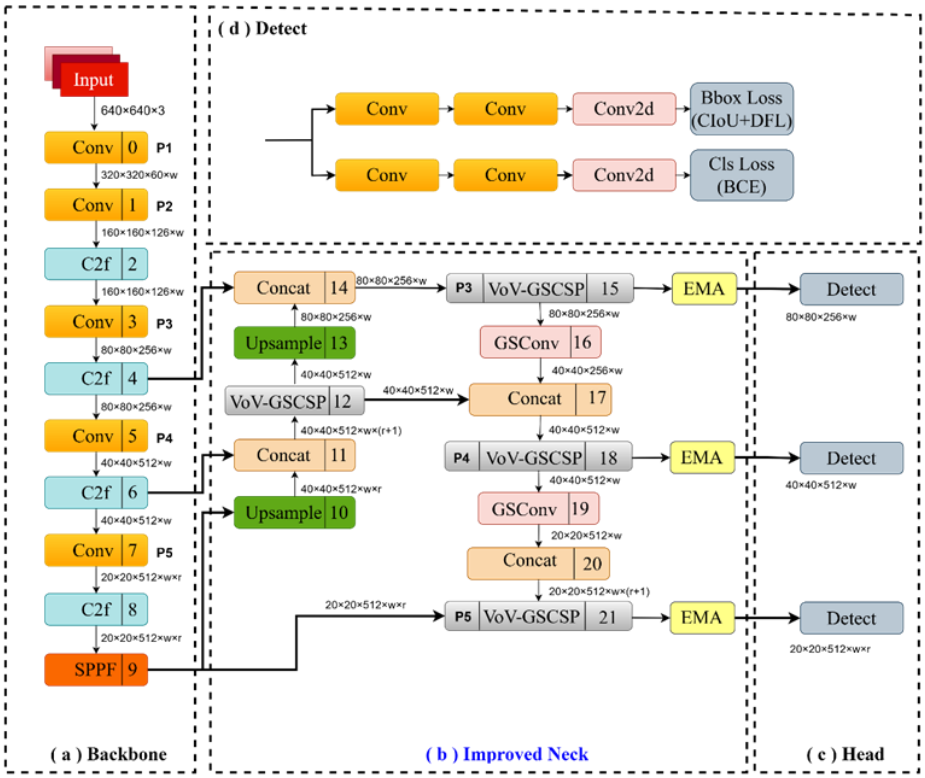

Figure 4. Structure of Modified YOLOv8n

Figure 4 outlines the modified network structure of the Modified YOLOv8n, highlighting the incorporation of lightweight mechanisms and attention modules within the neck structure. By replacing the original C2f modules with VoV-GSCSP blocks and integrating GSConv layers, the design achieves enhanced computational efficiency while retaining the network's feature representation capabilities. The inclusion of EMA modules after the P3, P4, and P5 layers strengthens the model's capacity to prioritize subtle and small-scale defect characteristics, particularly in visually cluttered and complex environments. These refinements significantly elevate detection accuracy without compromising real-time processing speeds, positioning the model as a robust solution for real-world inspection applications.

Vol. 7, No. 1, February 2026, Page. 60-80
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4679

2.4. Slim-Neck Structure with GSConv

Slim-neck is an efficient solution designed for the neck part of real-time object detection architectures, aiming to enhance accuracy while reducing computational burden. This approach is achieved through an innovative design that maintains high-quality feature representation while minimizing parameter count and computational complexity. One of its core components is the Ghost-Shuffle Convolution (GSConv) [33], which addresses the limitations of depthwise separable convolution (DSC). Although DSC effectively reduces parameters and computation, it suffers from poor inter-channel information exchange, which negatively impacts feature representation. GSConv combines the strengths of standard convolution (SC) and DSC through a channel-shuffling mechanism, enabling interaction across channels and preserving semantic connectivity. This hybrid approach approximates the feature representation quality of SC with approximately 50% less computation, while allowing better spatial and channel-level feature extraction. The core mechanism involves splitting input channels into groups, applying DSC independently to each group, and then shuffling and merging them, as depicted in Figure 5.

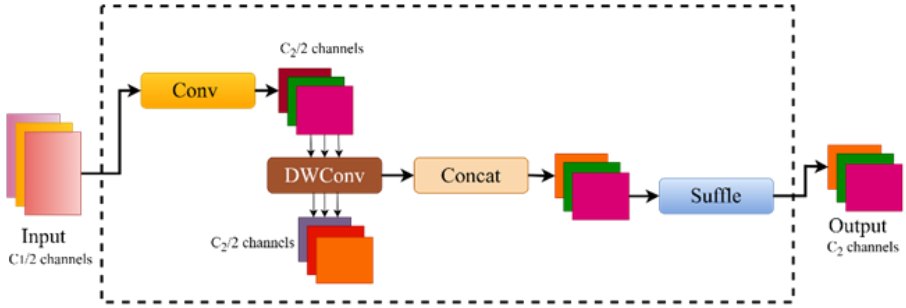

Figure 5. Ghost-Shuffle Convolution Module

Although GSConv has proven effective in reducing redundant information in feature maps while maintaining detection accuracy, it has limited impact on inference time. To address this limitation, the GSConv module was further developed through the integration of a GS bottleneck, as illustrated in Figure 6a, which replaces the default bottleneck in the C2f structure. This led to the development of a new module called VoV-GSCSP, a cross-layer feature refinement framework designed to improve multi-scale feature integration with greater computational efficiency. VoV-GSCSP replaces the ELAN module in the neck of the network and is constructed by combining functional blocks as shown in Figure 6b.

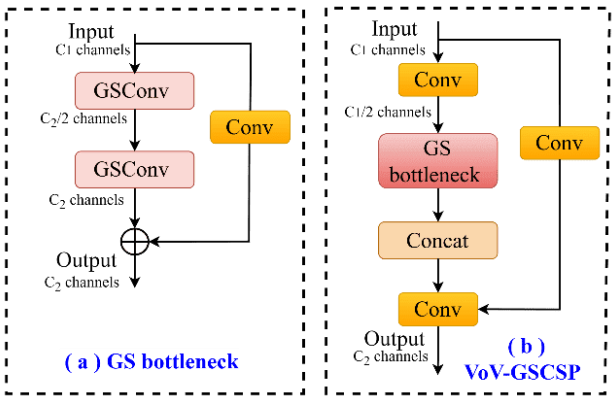

Figure 1. The architectural designs of (a) the GS bottleneck module and (b) the VoV-GSCSP module

Vol. 7, No. 1, February 2026, Page. 60-80
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4679

As depicted in Figure 6b, the workflow begins with a 1×1 convolutional layer that extracts initial features while reducing input channel dimensions. The output is then passed to the GS bottleneck, which utilizes a residual connection to enhance gradient flow. Within the bottleneck, data is processed through two successive GSConv layers to deepen spatial and channel-wise relationships. The refined features are then concatenated with the compressed output of a parallel 1×1 convolution (reduced to half the input channels). This merged output is again fused with the original VoV-GSCSP input using an additional 1×1 convolution, followed by a final convolutional layer that restores the number of output channels to C_2, as mathematically outlined in Equations (2) and (3).

$$\mathbf{GSB_{out}} = \mathbf{F_{GSC}}\left(\mathbf{F_{GSC}}(\boldsymbol{\alpha}\mathbf{X_{C1}})\right) + \boldsymbol{\alpha}(\mathbf{X})_{\mathbf{C1/2}} \quad (2)$$

$$\mathbf{VoVGSCSP_{out}} = \boldsymbol{\alpha}(\mathbf{Concat}(\mathbf{GSB_{out}}, (\boldsymbol{\alpha}\mathbf{X_{c1}}))) \quad (3)$$

2.5. Neck Network with EMA

Figure 4b schematically illustrates the architectural modifications to the neck network in the enhanced YOLOv8n model. The design integrates EMA modules [32] following the final VoV-GSCSP blocks at the P3, P4, and P5 stages, preserving the original spatial resolution of feature maps while enhancing semantic discriminability. As depicted in Figure 7, the EMA mechanism operates as a critical element within the enhanced architecture. For an input feature map $\boldsymbol{X} \in \mathbb{R}^{\boldsymbol{C}\times\boldsymbol{H}\times\boldsymbol{W}}$, the EMA module partitions the channels into G parallel sub-features ($\boldsymbol{X} = [\boldsymbol{X_0}, \boldsymbol{X_1}, \ldots, \boldsymbol{X_{G-1}}]$, where $\boldsymbol{X_i} \in \mathbb{R}^{\frac{\boldsymbol{C}}{\boldsymbol{G}}\times\boldsymbol{H}\times\boldsymbol{W}}$), facilitating multi-group learning of heterogeneous semantic patterns. This grouped processing enables the network to capture nuanced defect signatures across diverse spatial and contextual scales, particularly beneficial for small or occluded defects in complex environments.

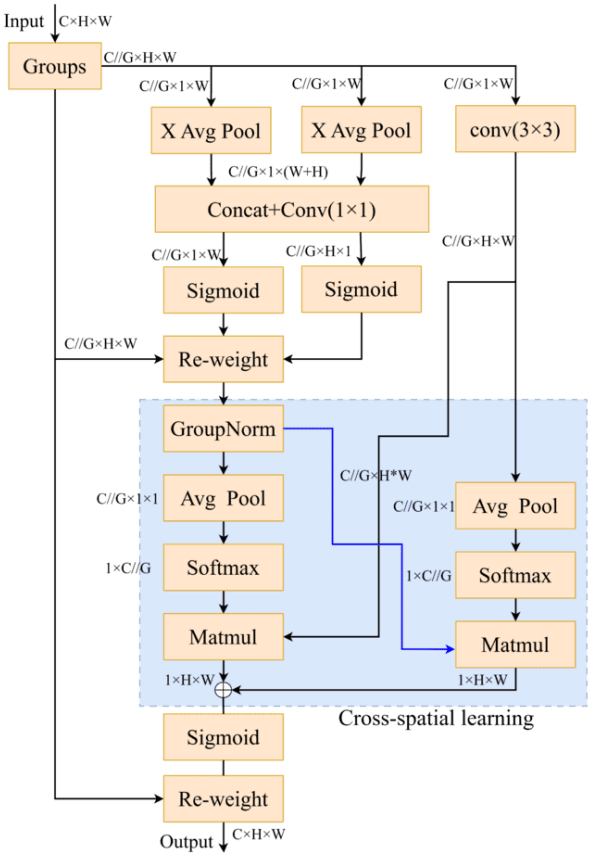

Figure 2. Structure of the EMA module

Vol. 7, No. 1, February 2026, Page. 60-80
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4679

To augment the model's capability for integrate multi-scale spatial information, the original 3×3 convolutional branch is substituted with a 5×5 branch, broadening the network's receptive field to capture richer contextual patterns across varying scales. The EMA module utilizes three parallel pathways to derive attention descriptors from grouped feature maps: two branches employ 1×1 convolutions for localized feature extraction, while the third incorporates a 5×5 convolution to encode broader spatial relationships. This design introduces a novel cross-dimensional aggregation strategy, enabling seamless fusion of global context and fine-grained local details. Specifically, the outputs of the 1×1 convolutional branches undergo 2D global average pooling to distill channel-wise global spatial statistics, condensing features into compact descriptors. Simultaneously, the output from the smaller-scale branch is reconfigured to meet the dimensional specifications necessary for the ensuing channel-wise feature activation step. The global pooling operation, mathematically expressed as:

$$z_c = \frac{1}{H.W}\sum_{i=1}^{H}\sum_{j=1}^{W} x_c(i,j) \quad (4)$$

aims to capture global context and model long-range spatial dependencies across each channel. This approach enables EMA to enhance the model's sensitivity to spatial relationships, particularly in complex visual environments.

3. EXPERIMENTAL RESULTS

This section outlines the experimental setup, including the datasets used, evaluation metrics, selected hyperparameters, as well as the results obtained and their corresponding analysis.

3.1. Experimental Environment and Datasets

The experimental environment runs on a 64-bit Windows 11 operating system, utilizing a GPU-enabled system to accelerate deep learning processes. The development of the model was carried out using the PyTorch framework with support from CUDA and Python. An overview of the hardware and software components used in this experiment is presented in Table I.

Table I. Hardware and Software Specifications

Description	Specification
Processor	Intel Core i7-10700F
GPU	Nvidia GeForce RTX 3050
RAM	32GB
Operating System	Windows 11
Software Libraries	CUDA 11.8, Python 3.8.5, PyTorch 2.5.1

The dataset used in this study is the public Insulator Defect Image Dataset (IDID) [34], which can be accessed through the IEEE Dataport platform. This dataset is specifically designed for defect detection tasks in power transmission systems, with a focus on images containing insulator strings captured in outdoor environments. The IDID includes insulators in three primary conditions: (1) flashover, where visible electrical discharge marks are present; (2) broken, referring to structural damage or cracks in the ceramic shell of the insulator; and (3) normal, which indicates insulators without any observable defects. The original dataset comprises 1,600 high-resolution images, each with sufficient detail to identify fine-grained visual features necessary for effective training of deep learning models. samples of the original dataset are presented in Figure 8 to provide visual context of the image quality and class representation.

Vol. 7, No. 1, February 2026, Page. 60-80
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4679

Figure 3. Image sample from IDID dataset

To ensure high-quality training data, all images were annotated manually using the Roboflow annotation tool. This platform provides an intuitive interface that facilitates precise bounding box placement around defective areas. To further enhance the model's ability to generalize to real-world deployment scenarios such as those encountered during drone-based inspection in outdoor environments an extensive data augmentation strategy was employed. The augmentation techniques applied include horizontal and vertical flipping, brightness adjustment, and Gaussian blur, each simulating common variations in image capture caused by camera movement, lighting conditions, and object orientation. These augmentations not only improve robustness but also address dataset imbalance by artificially increasing the number of training samples for minority classes.

The distribution of bounding box annotations per class in the augmented Insulator Defect Image Dataset (IDID) is summarized in Table II, which provides a quantitative overview of the dataset composition. As shown in the table, the "Insulator" class contains a total of 1,093 annotations, while defect-related classes such as "Broken" and "Flashover" are more prevalent, with 2,448 and 1,832 annotations, respectively. This distribution indicates a deliberate emphasis on capturing various defect scenarios, ensuring that the dataset is sufficiently representative for training robust deep learning models. Such a distribution supports the development of models that can generalize well across different defect types, especially considering the imbalanced yet realistic nature of field data. A higher frequency of certain defect classes also encourages the model to better distinguish subtle variations within similar visual patterns.

Furthermore, Figure 9 presents a sample of annotated images from the dataset, highlighting the wide range of visual conditions encountered during data collection. These samples demonstrate the challenges inherent in real-world insulator defect detection tasks, including partial, and complex, cluttered backgrounds. These factors introduce significant variation in object appearance and context, demanding that the detection model learns robust feature representations to maintain high accuracy in practical deployment scenarios.

Table II. Annotation Distribution per Class

Class	Total Annotations
Insulator	1093
Broken	2448
Flashover	1832

Vol. 7, No. 1, February 2026, Page. 60-80
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4679

Figure 4. Sample annotated images from the IDID dataset

3.2. Evaluation Metrics and Hyperparameters Settings

In object detection, model effectiveness is mainly assessed through Average Precision (AP), which captures the accuracy in both identifying and categorizing objects. When the task includes several object categories, the evaluation expands to mean Average Precision (mAP), calculated as the average AP across all classes. Alongside accuracy, the inference speed is a key consideration, particularly for real-time systems. This is quantified by Frames Per Second (FPS), which measures the number of images the model can analyze within one second. The mathematical formulations of these evaluation metrics are provided in Equations (5) to (9):

$$P = \frac{TP}{TP+FP}, \tag{5}$$

$$R = \frac{TP}{TP+FN}, \tag{6}$$

$$AP = \int_0^1 P(R)d\,R, \tag{7}$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \tag{8}$$

$$FPS = \frac{1000}{Total\ Processing\ Time\ per\ Image} \tag{9}$$

Where TP (True Positives) represents the number of correctly predicted positive samples, FN (False Negatives) refers to the number of actual positive samples that were incorrectly predicted as negative, and FP (False Positives) denotes the number of negative samples that were incorrectly classified as positive. The symbol R stands for Recall, P for Precision, N is the total number of object categories being detected, and Total processing time per image includes preprocessing, inference, and postprocessing times.

In this experiment, the model training process was configured with carefully selected hyperparameters to ensure convergence stability and efficient learning. Stochastic Gradient Descent (SGD) was employed as the optimization algorithm due to its robustness in large-scale learning tasks. The initial learning rate was set to 0.01, a balanced value chosen to prevent divergence from overly high rates and slow convergence from overly low rates. To further enhance training stability and generalization, a momentum of 0.95 and a weight decay of 0.0005 were applied. These values were determined based on iterative experimental trials and performance monitoring through visualized

Vol. 7, No. 1, February 2026, Page. 60-80
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4679

training graphs. Although the model showed signs of convergence after approximately 100 epochs, the training duration was extended to 300 epochs to ensure performance consistency across the dataset. A batch size of 32 was selected to accommodate the available computational resources, maintaining a balance between memory usage and gradient estimation stability. Other parameters, such as the input resolution and augmentation ratios, followed default configurations from the Ultralytics YOLO framework. A summary of the key hyperparameter settings used in the experiment is presented in Table III.

Table III. Hyperparameters Settings

Parameter	*Value*
Epochs	300
Batch Size	32
Optimizer	Stochastic Gradient Descent (SGD)
Initial Learning Rate	0.01
Weight Decay	0.0005
Momentum	0.937

3.3. Experimental results and analysis Experiments

To validate the effectiveness of the proposed method, a series of experiments and comparative analyses were conducted using the Insulator Defect Image Dataset (IDID). The evaluation process is divided into two main components: ablation studies and comparative experiments. (1) To thoroughly investigate the role of each modification introduced in the proposed approach, a series of ablation experiments were conducted. These experiments aim to isolate the impact of each component by evaluating model performance with and without specific enhancements. This process allows for a deeper understanding of how individual changes contribute to overall improvements in detection accuracy, computational efficiency, and robustness in diverse visual conditions. (2) In parallel, comparative experiments were performed to benchmark the proposed method against several existing object detection models based on the YOLO family. Using the same dataset and experimental settings, the proposed model was evaluated alongside baseline architectures.

3.4. Ablation Experiment

To comprehensively evaluate the impact of each architectural enhancement introduced in the proposed method, an ablation study was conducted focusing on two key components: the Efficient Multi-Scale Attention (EMA) module and the Slim-neck structure based on GSConv. These components were integrated separately and jointly into the YOLOv8n baseline to isolate and quantify their individual and combined contributions to the model's overall performance. The ablation experiments are essential to understand how each modification affects detection accuracy, model complexity, and inference speed. Table IV presents the detailed results of this analysis, comparing variations of the model with and without the proposed modules across key metrics, including class-wise Average Precision (AP), mean Average Precision at 0.5 IoU (mAP@0.5), FLOPs, parameter count, and Frames Per Second (FPS).

As shown in Table IV, the baseline YOLOv8n model, without any enhancements, achieved a mean Average Precision (mAP@0.5) of 90.5%, with class-wise APs of 95.5% for Broken insulators, 84.6% for Flashover, and 93.5% for general Insulator detection, operating at a speed of 178.57 frames per second (FPS). When the EMA module was introduced independently, a marginal improvement in overall detection accuracy was observed, with mAP@0.5 increasing to 90.8%. Notably, the AP for Flashover defects increased to 85.3%, indicating that EMA enhances the model's ability to capture nuanced spatial features, particularly in defect classes with complex visual patterns and subtle luminance

Vol. 7, No. 1, February 2026, Page. 60-80
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4679

transitions. However, a slight reduction in AP for Broken defects suggests that the added spatial aggregation could lead to an over-smoothing effect, potentially reducing sensitivity to fine-grained edge discontinuities present in fracture patterns. Additionally, the computational cost slightly increased (FLOPs: 8.2G), and inference speed dropped to 158.73 FPS, underscoring the trade-off between spatial precision and processing efficiency.

Table IV. Results of the ablation experiments

Model	Slim-neck	EMA	AP(%)			mAP	FLOPs	Params	FPS
			Broken	Flashover	Insulator	@0.5(%)	(G)	(M)	
YOLOv8n	-	-	95.5	84.6	93.5	90.5	8.1	3	178.57
	-	✓	93.4	85.3	93.7	90.8	8.2	2.9	158.73
	✓	-	94.5	85.7	93.3	91.2	7.3	2.7	185.19
	✓	✓	95.6	86.9	94	92.1	7.4	2.7	161.29

In contrast, replacing the standard neck structure with a lightweight Slim-neck alone yielded a more notable enhancement in both accuracy and efficiency. The model attained a mAP of 91.2% with a significant reduction in FLOPs to 7.3G and parameter count to 2.7M, achieving the highest FPS of 185.19 across all configurations. These results suggest that Slim-neck not only accelerates inference but also preserves semantic richness across scales due to its efficient feature aggregation strategy beneficial in processing aerial images with small object-to-background ratios, as commonly encountered in power grid inspection. Crucially, the integration of both EMA and Slim-neck resulted in the best performance, achieving a peak mAP@0.5 of 92.1%, with respective APs of 95.6%, 86.9%, and 94.0 for Broken, Flashover, and Insulator classes. While the inclusion of EMA slightly increases the FLOPs to 7.4G and reduces the speed to 161.29 FPS compared to Slim-neck alone, the combined configuration offers the optimal trade-off between accuracy and efficiency. The boost in performance is attributed to the complementary nature of the two modules: Slim-neck enhances feature propagation and compactness, while EMA facilitates robust contextual modeling by focusing attention on the most discriminative spatial regions.

This synergy is visually corroborated in Figure 10. In the top row, the model with EMA correctly identifies subtle fractures that were either missed or poorly localized by the baseline YOLOv8n. The bounding boxes are more precise, and confidence scores are noticeably higher, indicating greater certainty and improved spatial focus. In the second row, for Flashover defects, the EMA-augmented model suppresses false positives and demonstrates more stable label assignments, even under complex background textures. Furthermore, the EMA-equipped model demonstrates a clear advantage in distinguishing between overlapping or adjacent insulator components a common challenge in UAV-based inspection tasks where perspective distortion and occlusion are prevalent. These findings validate that the proposed enhancements substantially elevate the model's ability to detect minute and complex defects with higher fidelity, while still maintaining high throughput. More importantly, the reduction in computational overhead via Slim-neck ensures that the model remains deployable on edge devices, making it suitable for real-time aerial inspection scenarios. Therefore, the final configuration combining Slim-neck and EMA presents an optimal balance of detection precision, computational complexity, and real-time capability addressing critical challenges in small object detection within high-resolution UAV imagery.

Vol. 7, No. 1, February 2026, Page. 60-80
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4679

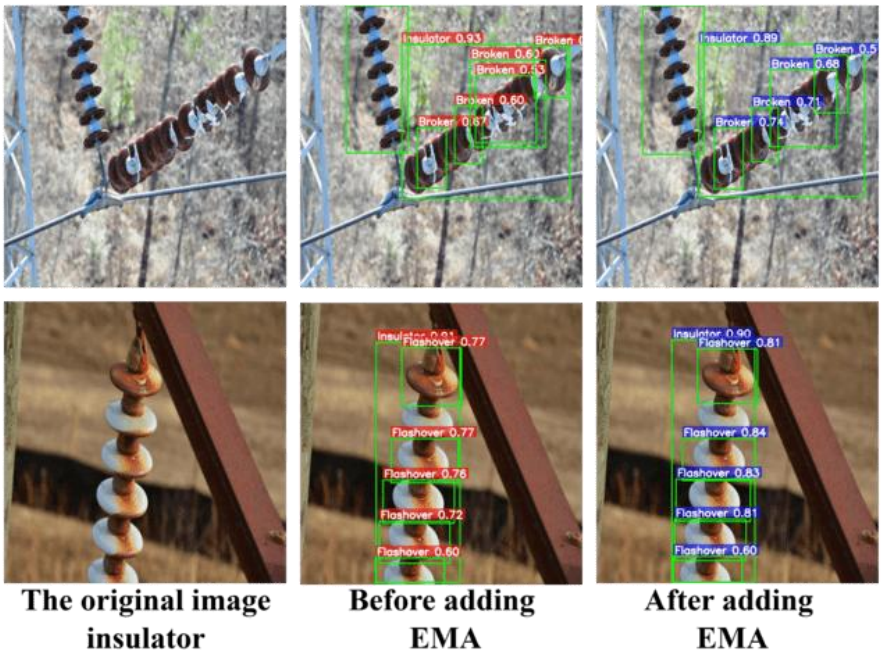

Figure 5. Comparison before and after adding EMA attention mechanism on the IDID dataset.

3.5. Comparison Experiment

The ablation experimental results show that the integration of the Slim-neck structure and the Efficient Multi-Scale Attention (EMA) module significantly improves the performance of the YOLOv8n baseline model in terms of both detection accuracy and computational efficiency. These findings confirm that each component contributes positively, with their combination yielding the most balanced and effective model configuration. To further validate the general effectiveness and competitiveness of the proposed model, this section presents a comparison experiment against several widely used YOLO-based object detection algorithms, including YOLOv5n, YOLOv6n, and the original YOLOv8n. The purpose of this experiment is to benchmark the proposed method's performance using the same dataset and evaluation metrics, thereby assessing its practical advantages in real-world defect detection tasks. The comparative results, summarized in Table V, highlight key indicators such as precision, recall, mean Average Precision (mAP), computational cost, and inference speed, offering a comprehensive view of the model's capabilities relative to other existing methods.

Table V. Performance comparison with other models on the IDID dataset.

Model	P (%)	R (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)	Params (M)	FLOPs (G)	FPS
YOLOv5n	89.4	87.6	91.1	73.5	2.5	7.1	178.57
YOLOv6n	89.3	84.6	89	74.9	4.2	11.9	185.19
YOLOv8n	90.1	88.5	90.5	74.8	3	8.1	178.57
Proposed Method	**92**	**88.6**	**92.1**	**75.1**	**2.7**	**7.4**	**161.29**

The proposed method achieves the highest Precision score of 92.0%, demonstrating a superior ability to correctly identify positive instances compared to YOLOv5n with 89.4%, YOLOv6n with 89.3%, and YOLOv8n with 90.1%. In terms of Recall, the proposed model also leads with 88.6%, slightly ahead of YOLOv8n at 88.5%, and significantly better than YOLOv5n at 87.6% and YOLOv6n at 84.6%. These results indicate strong robustness of the proposed method in identifying true defect

Vol. 7, No. 1, February 2026, Page. 60-80
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4679

instances, even under complex visual scenarios. For localization performance, the model attains the highest mAP@0.5 at 92.1%, surpassing YOLOv5n with 91.1%, YOLOv6n with 89.0%, and YOLOv8n with 90.5%. Additionally, in the more rigorous mAP@0.5:0.95 evaluation, the proposed model achieves 75.1%, slightly higher than YOLOv5n at 73.5%, YOLOv6n at 74.9%, and YOLOv8n at 74.8%. This confirms that the proposed model not only performs well on standard IoU thresholds but also maintains accuracy across a wide range of object localization conditions. In terms of computational efficiency, the proposed method is notably lightweight with only 2.7 million parameters and 7.4 billion FLOPs. This is more efficient than YOLOv6n with 4.2 million parameters and 11.9 billion FLOPs, and YOLOv8n with 3.0 million parameters and 8.1 billion FLOPs. Despite this compact structure, the model does not compromise on accuracy, highlighting the effectiveness of the Slim-neck design and GSConv integration. The only trade-off observed is in the inference speed. The proposed model operates at 161.29 FPS, which is slightly lower than YOLOv5n at 178.57 FPS and YOLOv6n at 185.19 FPS. This decrease is attributed to the addition of the Efficient Multi-Scale Attention module, which introduces additional processing to enhance feature representation. However, the drop in speed remains within real-time limits and is justified by the gains in accuracy and efficiency.

To provide a more comprehensive evaluation beyond numerical performance metrics, Figures 11 and 12 present qualitative comparisons of the proposed model's detection capability under various conditions. These visualizations serve as essential complements to the quantitative analysis by illustrating how the model behaves in real-world scenarios, where factors such as image quality, environmental complexity, and lighting variability play significant roles in object detection performance. Figure 11 presents a comparative evaluation of detection performance between the proposed model and three baseline YOLO variants: YOLOv5n, YOLOv6n, and YOLOv8n. The figure is structured in three horizontal panels, each representing a distinct scene involving different insulator defect types, and four vertical columns representing detection results from the respective models. This visualization enables a clear assessment of how each model performs in terms of localization accuracy, detection confidence, and visual reliability under real-world scenarios.

In the first row of images, which features insulator chains mounted on a metallic tower structure, the detection of Broken components proves to be challenging due to repetitive patterns and the presence of structural interference. YOLOv5n and YOLOv6n produce results with relatively low confidence, such as 0.64 and 0.70 for Broken defects, and often draw bounding boxes that are either misaligned or overly constrained. YOLOv8n shows moderate improvement in both box placement and confidence. However, the proposed model clearly outperforms the baselines by generating precise bounding boxes with confidence scores reaching 0.95 for Insulator and 0.71 for Broken. The localization is more accurate and better fitted to the actual defect area, indicating that the model is able to effectively extract fine-grained spatial features, even when surrounded by metallic noise and shadow artifacts.

In the second row, where flashover defects appear along the surfaces of stacked disc insulators with a grassy or textured background, the baseline models show varying degrees of overprediction and misclassification. YOLOv5n assigns multiple overlapping bounding boxes with low confidence values such as 0.40 and 0.51, suggesting poor defect discrimination. YOLOv6n performs slightly better, with a stronger confidence level of 0.66, but still suffers from visual redundancy. YOLOv8n improves the precision further, yet still exhibits a degree of label dispersion and inconsistent box alignment. In comparison, the proposed model generates clean, well-placed bounding boxes with confidence values reaching up to 0.60, 0.56, and 0.44 for Flashover, while simultaneously identifying the Insulator class with a high confidence of 0.93. These results indicate that the proposed model is more capable of isolating actual defect regions and suppressing irrelevant visual information, which is particularly crucial in outdoor scenes where background textures can easily lead to false detections.

Vol. 7, No. 1, February 2026, Page. 60-80
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4679

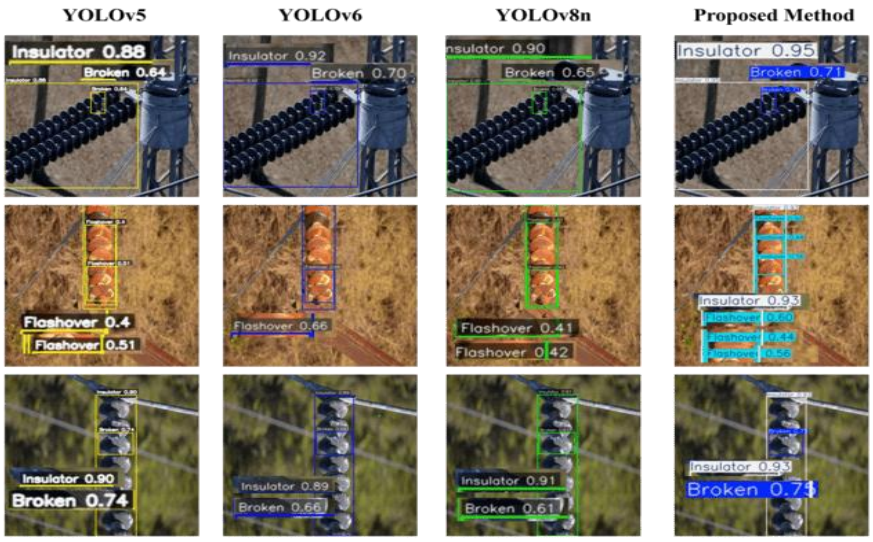

Figure 6. Comparison of detection result with other models.

The third row of Figure 11 displays insulators captured in visually degraded conditions, such as blurry backgrounds, lower contrast, and complex natural scenery. In this scenario, YOLOv5n and YOLOv6n manage to detect Broken defects with confidence around 0.74 and 0.66 but still exhibit instability in bounding box accuracy and class distinction. YOLOv8n performs relatively better but still underperforms in terms of certainty. The proposed model not only matches or exceeds the confidence of prior variants reaching 0.75 for Broken but also offers improved alignment between the predicted boxes and actual defect boundaries. It distinguishes Insulator and Broken classes with greater clarity, even under subtle image noise and low-contrast areas. Across all cases in Figure 11, it is evident that the proposed model consistently surpasses YOLOv5n, YOLOv6n, and YOLOv8n in terms of detection quality, confidence reliability, and robustness in spatial reasoning.

Figure 12 presents a comprehensive visual comparison that emphasizes the robustness of the proposed model over the YOLOv8n baseline under challenging real-world conditions commonly encountered during aerial inspections. In Row A, which captures conditions involving motion blur and defocus caused by UAV movement or unstable camera focus, the proposed model demonstrates superior stability in both detection accuracy and bounding box precision. It consistently identifies critical objects such as flashovers and insulators with high confidence, with examples reaching confidence scores as high as 0.92 for insulators and 0.66 for flashovers. Meanwhile, the YOLOv8n baseline fails to maintain detection consistency, exhibiting reduced confidence and occasionally missing targets altogether. This performance gap highlights the effectiveness of the proposed enhancements in preserving spatial understanding, even when low-level visual cues are degraded.

In Row B, which represents images with visually complex backgrounds including vegetation, towers, and metallic infrastructure, the proposed model again outperforms the baseline by maintaining precise localization of defects while avoiding false positives. The baseline model frequently misclassifies background textures and edges as defect classes, suggesting a weak contextual understanding. In contrast, the proposed model accurately isolates target objects like flashover regions, producing predictions with strong confidence values such as 0.72, and avoids confusing similar-looking regions in the environment. This improvement is largely attributed to the model's enhanced spatial attention and feature aggregation capability, which enables it to distinguish between foreground anomalies and background clutter an essential feature for deployment in visually dense inspection scenes.

Vol. 7, No. 1, February 2026, Page. 60-80
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4679

In Row C, which illustrates scenarios under extreme lighting conditions such as overexposure and glare from direct sunlight, the baseline YOLOv8n model shows clear limitations. It often produces weak or inconsistent predictions, failing to detect broken components reliably due to reduced contrast and visual detail. The proposed model, on the other hand, continues to deliver stable detections even under these harsh lighting conditions. It maintains high confidence scores, including values reaching up to 0.76 for broken insulators, reflecting a stronger photometric invariance and a more robust abstraction of object shape and structure. This ability to perform under challenging illumination is critical in outdoor inspections, where lighting variability is inevitable and uncontrolled.

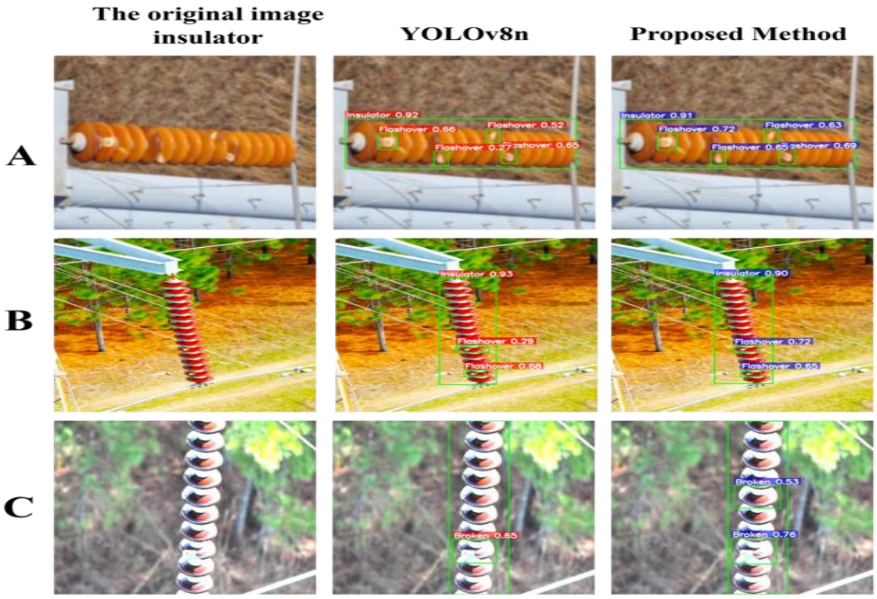

Figure 7. Comparison of detection results between baseline and proposed models under challenging visual conditions.

4. DISCUSSIONS

This study presents an improved object detection model by integrating a Slim-Neck structure with Efficient Multi-scale Attention (EMA) into the YOLOv8n architecture. The modified model achieves significant performance gains, increasing mAP@0.5 from 90.5% to 92.1%, while also reducing model size and computational complexity. The number of parameters decreases by 10.8%, from 3.0 million to 2.7 million, and FLOPs are reduced by 8.6%, from 8.1 billion to 7.4 billion. The model maintains a high inference speed of 161 FPS, making it practical for real-time UAV-based inspections. These improvements are driven by the use of GSConv, which retains rich channel interactions at lower computational cost, and the EMA module, which expands the receptive field and focuses on the most discriminative spatial and channel-wise features. Together, these enhancements improve performance under challenging conditions such as motion blur, complex backgrounds, and small object sizes.

Ablation studies reveal the individual contributions of each component. Adding EMA alone improves detection for flashover defects but slightly reduces accuracy for broken insulators, likely due to over-smoothing. The Slim-Neck module alone achieves the highest inference speed at 185.2 FPS and enhances multi-scale feature distribution. When combined, the Slim-Neck and EMA deliver the best performance, achieving a mAP@0.5 of 92.1% while maintaining low computational cost. These findings indicate that Slim-Neck improves gradient flow and preserves spatial information, particularly beneficial for detecting small objects. Meanwhile, EMA introduces contextual awareness, enhancing confidence scores in images with complex or noisy visual patterns.

Vol. 7, No. 1, February 2026, Page. 60-80
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4679

When compared to baseline models such as YOLOv5n, YOLOv6n, and YOLOv8n, the proposed approach achieves the highest performance in both mAP and precision, while maintaining one of the lightest model sizes. Beyond outperforming these standard models, it also remains highly competitive against more recent state-of-the-art approaches. For instance, the Insu-YOLO model by Y. Chen et al. [35], which is based on YOLOv8n and incorporates GSConv and CARAFE, demonstrates improved detection for small objects. However, its performance is limited when dealing with extremely small defects and it lacks refinement in the post-processing stage. Similarly, Su et al. [25] introduced Triplet Attention and a GSConv-based Slim-Neck into YOLOv8n to reduce background interference, but this came at the cost of lower precision and reduced inference speed. Meanwhile, the IALF-YOLO model proposed by Mei et al. [36] enhances small object detection by employing S-CBAM and GSConv, yet results in a more complex architecture that is challenging to deploy on embedded platforms. In contrast, the proposed model offers a more effective trade-off between detection accuracy and computational efficiency. By integrating a Slim-Neck module based on Ghost-Shuffle Convolution and an Efficient Multi-Scale Attention (EMA) mechanism, it achieves a precision of 92.0%, a mAP@0.5 of 92.1%, and a real-time inference speed of 161 FPS, while reducing the parameter count by 10.8% and computational load by 8.6%. These results demonstrate that the proposed method not only surpasses previous approaches in accuracy but also provides practical advantages for deployment in real-world, resource-constrained environments.

This research contributes to several key areas within the field of computer science. In computer vision and embedded AI, it demonstrates how attention mechanisms can be effectively integrated into lightweight architectures for real-time deployment. In the area of automated inspection, the model supports the use of UAVs for efficient power grid monitoring, advancing the application of AI for Social Good. In model design, it introduces a hybrid structure that combines structural re-parameterization, offering a practical design pattern for object detection in complex environments. Additionally, this study contributes to dataset analysis by demonstrating how techniques like augmentation and class-weighting can mitigate challenges such as class imbalance and image noise, thus supporting broader adoption in real-world computer vision tasks. This work advances the application of efficient deep learning models in real-time fault detection for smart grid infrastructures, providing a foundation for future intelligent systems capable of operating reliably under harsh and variable field conditions. Through this contribution, the study not only enhances technical capability but also strengthens the intersection between AI innovation and its deployment in critical societal infrastructure.

5. CONCLUSION

This study introduced an enhanced object detection model for insulator defect detection, built upon the YOLOv8n architecture and improved through the integration of a Slim-Neck structure and Efficient Multi-scale Attention (EMA). The proposed approach demonstrated significant gains in accuracy, achieving a mAP@0.5 of 92.1%, while maintaining a lightweight structure with 2.7 million parameters and an inference speed of 161 FPS. These improvements highlight the model's ability to perform real-time detection efficiently, making it well-suited for UAV-based inspection tasks in power transmission systems. By combining GSConv for structural re-parameterization with EMA for better contextual understanding, the model effectively addresses common challenges in aerial imagery, such as small object sizes, motion blur, and complex backgrounds. In addition to its strong technical performance, this research contributes to the broader fields of computer vision and embedded AI by demonstrating how attention mechanisms can be applied to create accurate, fast, and resource-efficient detection systems. The findings support practical applications in infrastructure monitoring and offer a reference for lightweight yet powerful model design.

Vol. 7, No. 1, February 2026, Page. 60-80
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4679

Looking ahead, the researcher hopes that this work can be further developed to build a more comprehensive and deployable insulator defect detection system. One key direction for future research is real-time testing using UAVs to capture live video streams in operational environments. Such testing would enable evaluation of the model's responsiveness, stability, and accuracy under field conditions. Furthermore, deploying the model on edge devices such as Jetson Nano or Raspberry Pi is essential for assessing its performance in resource-constrained settings, and for ensuring robustness in practical deployments. Another critical recommendation is to expand and diversify the dataset by including more types of insulator defects, various weather conditions (e.g., sunny, rainy, foggy), and different image perspectives or camera angles. These additions would significantly enhance the model's generalization capabilities and make it more adaptable to the variability of real-world inspection scenarios. In summary, this research lays a solid foundation for AI-based infrastructure monitoring, and its continued development through real-world testing and dataset expansion will be vital for advancing intelligent, scalable inspection systems in the field of computer science.

CONFLICT OF INTEREST

There is no potential conflict of interesting this work.

REFERENCES

[1] A. Odo, S. McKenna, D. Flynn, and J. B. Vorstius, "Aerial Image Analysis Using Deep Learning for Electrical Overhead Line Network Asset Management," *IEEE Access*, vol. 9, pp. 146281–146295, 2021, doi: 10.1109/ACCESS.2021.3123158.

[2] H. Liang, C. Zuo, and W. Wei, "Detection and Evaluation Method of Transmission Line Defects Based on Deep Learning," *IEEE Access*, vol. 8, pp. 38448–38458, 2020, doi: 10.1109/ACCESS.2020.2974798.

[3] Z. Lu, Y. Li, F. Shuang, and C. Han, "InsDef: Few-Shot Learning-Based Insulator Defect Detection Algorithm With a Dual-Guide Attention Mechanism and Multiple Label Consistency Constraints," *IEEE Trans. Power Deliv.*, vol. 38, no. 6, pp. 4166–4178, Dec. 2023, doi: 10.1109/TPWRD.2023.3311643.

[4] W. Rodgers, J. A. Cardenas, L. A. Gemoets, and R. J. Sarfi, "A smart grids knowledge transfer paradigm supported by experts' throughput modeling artificial intelligence algorithmic processes," *Technol. Forecast. Soc. Change*, vol. 190, p. 122373, May 2023, doi: 10.1016/j.techfore.2023.122373.

[5] Y. Cao, H. Xu, C. Su, and Q. Yang, "Accurate Glass Insulators Defect Detection in Power Transmission Grids Using Aerial Image Augmentation," *IEEE Trans. Power Deliv.*, vol. 38, no. 2, pp. 956–965, Apr. 2023, doi: 10.1109/TPWRD.2022.3202958.

[6] F. Alcayde-García, E. Salmerón-Manzano, M. A. Montero, A. Alcayde, and F. Manzano-Agugliaro, "Power Transmission Lines: Worldwide Research Trends," *Energies*, vol. 15, no. 16, p. 5777, Aug. 2022, doi: 10.3390/en15165777.

[7] L. Zhang, B. Li, Y. Cui, Y. Lai, and J. Gao, "Research on improved YOLOv8 algorithm for insulator defect detection," *J. Real-Time Image Process.*, vol. 21, no. 1, pp. 1–14, 2024, doi: 10.1007/s11554-023-01401-9.

[8] X. Liu, X. Miao, H. Jiang, and J. Chen, "Data analysis in visual power line inspection: An in-depth review of deep learning for component detection and fault diagnosis," *Annu. Rev. Control*, vol. 50, pp. 253–277, 2020, doi: 10.1016/j.arcontrol.2020.09.002.

[9] A. El-Hag, "Application of Machine Learning in Outdoor Insulators Condition Monitoring and Diagnostics," *IEEE Instrum. Meas. Mag.*, vol. 24, no. 2, pp. 101–108, Apr. 2021, doi: 10.1109/MIM.2021.9400959.

[10] B. Chen, "Fault Statistics and Analysis of 220-kV and Above Transmission Lines in a Southern Coastal Provincial Power Grid of China," *IEEE Open Access J. Power Energy*, vol. 7, pp. 122–129, 2020, doi: 10.1109/OAJPE.2020.2975665.

[11] Z. Wang, Q. Gao, J. Xu, and D. Li, "A Review of UAV Power Line Inspection," 2022, pp. 3147–

Vol. 7, No. 1, February 2026, Page. 60-80
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4679

3159. doi: 10.1007/978-981-15-8155-7_263.

[12] J. Hu, W. Wan, P. Qiao, Y. Zhou, and A. Ouyang, "Power Insulator Defect Detection Method Based on Enhanced YOLOV7 for Aerial Inspection," *Electronics*, vol. 14, no. 3, p. 408, Jan. 2025, doi: 10.3390/electronics14030408.

[13] F. FJ, L. Meganathan, M. Esha, N. G. Minh Thao, and A. S. Arockia Doss, "Power Transmission Line Inspection Using Unmanned Aerial Vehicle - A Review," in *2023 Innovations in Power and Advanced Computing Technologies (i-PACT)*, IEEE, Dec. 2023, pp. 1–5. doi: 10.1109/i-PACT58649.2023.10434671.

[14] J. Ding *et al.*, "Object Detection in Aerial Images: A Large-Scale Benchmark and Challenges," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 7778–7796, Nov. 2022, doi: 10.1109/TPAMI.2021.3117983.

[15] J. Liu, C. Liu, Y. Wu, H. Xu, and Z. Sun, "An Improved Method Based on Deep Learning for Insulator Fault Detection in Diverse Aerial Images," *Energies*, vol. 14, no. 14, p. 4365, Jul. 2021, doi: 10.3390/en14144365.

[16] Y. Liu, D. Liu, X. Huang, and C. Li, "Insulator defect detection with deep learning: A survey," *IET Gener. Transm. Distrib.*, vol. 17, no. 16, pp. 3541–3558, Aug. 2023, doi: 10.1049/gtd2.12916.

[17] Y. Gong *et al.*, "Defect detection of small cotter pins in electric power transmission system from UAV images using deep learning techniques," *Electr. Eng.*, vol. 105, no. 2, pp. 1251–1266, Apr. 2023, doi: 10.1007/s00202-022-01729-8.

[18] J. Liu, M. Hu, J. Dong, and X. Lu, "Summary of insulator defect detection based on deep learning," *Electr. Power Syst. Res.*, vol. 224, p. 109688, Nov. 2023, doi: 10.1016/j.epsr.2023.109688.

[19] X. Lu, Q. Li, B. Li, and J. Yan, "MimicDet: Bridging the Gap Between One-Stage and Two-Stage Object Detection," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12359 LNCS, pp. 541–557, 2020, doi: 10.1007/978-3-030-58568-6_32.

[20] H. Zhang and R. S. Cloutier, "Review on One-Stage Object Detection Based on Deep Learning," *EAI Endorsed Trans. e-Learning*, vol. 7, no. 23, p. 174181, Jun. 2022, doi: 10.4108/eai.9-6-2022.174181.

[21] B. Li, M. Xu, Z. Xie, D. Qi, and Y. Yan, "Research on Insulator Defect Detection Based on Improved YOLOv7," *Lect. Notes Electr. Eng.*, vol. 1169 LNEE, pp. 173–180, 2024, doi: 10.1007/978-981-97-1072-0_17.

[22] N. Zhang, J. Su, Y. Zhao, and H. Chen, "Insulator-YOLO: Transmission Line Insulator Risk Identification Based on Improved YOLOv5," *Processes*, vol. 12, no. 11, p. 2552, Nov. 2024, doi: 10.3390/pr12112552.

[23] L. Ding, Z. Q. Rao, B. Ding, and S. J. Li, "Research on Defect Detection Method of Railway Transmission Line Insulators Based on GC-YOLO," *IEEE Access*, vol. 11, pp. 102635–102642, 2023, doi: 10.1109/ACCESS.2023.3316266.

[24] Y. Zhang, J. Li, W. Fu, J. Ma, and G. Wang, "A lightweight YOLOv7 insulator defect detection algorithm based on DSC-SE," *PLoS One*, vol. 18, no. 12, p. e0289162, Dec. 2023, doi: 10.1371/journal.pone.0289162.

[25] J. Su, Y. Yuan, K. Przystupa, and O. Kochan, "Insulator defect detection algorithm based on improved YOLOv8 for electric power," *Signal, Image Video Process.*, vol. 18, no. 8–9, pp. 6197–6209, 2024, doi: 10.1007/s11760-024-03307-w.

[26] S. Luan, C. Li, P. Xu, Y. Huang, and X. Wang, "MI-YOLO: more information based YOLO for insulator defect detection," *J. Electron. Imaging*, vol. 32, no. 04, Jul. 2023, doi: 10.1117/1.JEI.32.4.043014.

[27] D. Wei, B. Hu, C. Shan, and H. Liu, "Insulator defect detection based on improved Yolov5s," *Front. Earth Sci.*, vol. 11, Feb. 2024, doi: 10.3389/feart.2023.1337982.

[28] Z. Li, C. Jiang, and Z. Li, "An Insulator Location and Defect Detection Method Based on Improved YOLOv8," *IEEE Access*, vol. 12, pp. 106781–106792, 2024, doi: 10.1109/ACCESS.2024.3436919.

[29] W. Han, Z. Cai, X. Li, A. Ding, Y. Zou, and T. Wang, "LMD-YOLO: A lightweight algorithm

Vol. 7, No. 1, February 2026, Page. 60-80
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.4679

for multi-defect detection of power distribution network insulators based on an improved YOLOv8," *PLoS One*, vol. 20, no. 2, p. e0314225, Feb. 2025, doi: 10.1371/journal.pone.0314225.

[30] Z. Mei, H. Xu, L. Yan, and K. Wang, "IALF-YOLO: Insulator defect detection method combining improved attention mechanism and lightweight feature fusion network," *Measurement*, vol. 253, p. 117701, Sep. 2025, doi: 10.1016/j.measurement.2025.117701.

[31] Z. Xu and X. Tang, "Transmission line insulator defect detection algorithm based on MAP-YOLOv8," *Sci. Rep.*, vol. 15, no. 1, p. 10288, Mar. 2025, doi: 10.1038/s41598-025-92445-3.

[32] D. Ouyang *et al.*, "Efficient Multi-Scale Attention Module with Cross-Spatial Learning," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, Jun. 2023, pp. 1–5. doi: 10.1109/ICASSP49357.2023.10096516.

[33] H. Li, J. Li, H. Wei, Z. Liu, Z. Zhan, and Q. Ren, "Slim-neck by GSConv: a lightweight-design for real-time detector architectures," *J. Real-Time Image Process.*, vol. 21, no. 3, pp. 1–11, 2024, doi: 10.1007/s11554-024-01436-6.

[34] P. K. Dexter Lewis, "Insulator Defect Detection," *IEEE Dataport*, 2021, doi: https://dx.doi.org/10.21227/vkdw-x769.

[35] Y. Chen, H. Liu, J. Chen, J. Hu, and E. Zheng, "Insu-YOLO: An Insulator Defect Detection Algorithm Based on Multiscale Feature Fusion," *Electronics*, vol. 12, no. 15, p. 3210, Jul. 2023, doi: 10.3390/electronics12153210.

[36] Z. Mei, H. Xu, L. Yan, and K. Wang, "IALF-YOLO: Insulator defect detection method combining improved attention mechanism and lightweight feature fusion network," *Meas. J. Int. Meas. Confed.*, vol. 253, no. PC, p. 117701, 2025, doi: 10.1016/j.measurement.2025.117701.