An Improved YOLOv11n Algorithm with Conv2Former and PW-IoU for UAV Inspection of Power Line Insulators

Bin Zheng
Faculty of Engineering, Mahasarakham University, Maha Sarakham, Thailand
65010363004@msu.ac.th

Niwat Angkawisittpan
Research Unit for Electrical and Computer Engineering Technology (RECENT), Mahasarakham University, Maha Sarakham, Thailand
niwat.a@msu.ac.th (corresponding author)

Lu Huang
Faculty of Electrical Engineering, Hunan Mechanical & Electrical Polytechnic, Changsha, Hunan, China
12936334@qq.com

Somchat Sonasang
Faculty of Industrial Technology, Nakhon Phanom University, Nakhon Phanom, Thailand
somchat.s@npu.ac.th

Received: 21 August 2025 | Revised: 8 October 2025 and 18 October 2025 | Accepted: 21 October 2025

 DOI: https://doi.org/10.48084/etasr.14220

ABSTRACT

Detecting insulator defects accurately and efficiently is vital for maintaining the reliability of power transmission systems, particularly during Unmanned Aerial Vehicle (UAV)-based inspections. In order to improve local and global feature extraction and detect minor, low-visibility flaws in complex settings, this paper suggests C2F-YOLOv11n, a lightweight detection framework that incorporates the Conv2Former attention mechanism. Experiments on a self-built insulator dataset show that C2F-YOLOv11n achieves 91.7% precision, 83.1% recall, 89.4% mAP50, and 58.9% mAP50-95, with inference speed at 194 FPS and a compact 2.70 MB model size, outperforming YOLOv8n and YOLOv10n. Additionally, a novel regression loss function, PW-IoU, combining PIoUv2's boundary-aware localization and WIoUv3's adaptive gradient reweighting, is introduced to address bounding box regression challenges. By integrating PW-IoU, the model achieves higher performance with precision reaching 92.6%, recall at 87.5%, mAP50 at 90.8%, and mAP50-95 at 59.3%, outperforming conventional Complete IoU (CIoU) and other IoU-related loss functions. PW-IoU enhances localization accuracy and convergence stability, especially for small targets in complex backgrounds. Furthermore, comparative experiments on the publicly available Chinese Power Line Insulator Dataset (CPLID) confirm the model's strong generalization, achieving competitive detection performance on both normal and defective insulators.

Keywords-Conv2Former; PW-IoU; insulator defect detection; small target detection

I. INTRODUCTION

Implementing effective and dependable inspection techniques has become crucial for ensuring the safe and stable functioning of power systems due to the power industry's rapid innovation and the growing proliferation of electrical equipment [1]. Insulators, usually placed between conductors with various electrical potentials or between conductors and grounding elements, are crucial parts of transmission lines. They are specifically designed to withstand both electrical stress and mechanical loads, thereby fulfilling the dual functions of supporting conductors and providing electrical insulation [2]. However, insulators are prone to issues such as flashover and fracture due to prolonged exposure to uncontrollable external environmental factors such as strong winds, hail, and freezing rain during long-term operation. In severe cases, these problems may threaten the safe and stable operation of the power grid [3]. Therefore, regular inspections

of insulators are of great importance to ensure the reliability of the power system and the continuity of power supply.

A range of conventional approaches has historically been applied to insulator defect identification, such as manual visual checks, acoustic-based diagnostics, thermal infrared scanning, and contactless radar measurement techniques. Conventional manual inspections predominantly rely on handheld detection tools, which are inefficient and pose safety risks [4]. With the continuous progress of Unmanned Aerial Vehicles (UAVs) and computer vision, the development of automated inspection systems has been substantially enhanced. UAVs equipped with high-resolution optical sensors have become the mainstream approach for intelligent inspection of insulators along high-voltage transmission lines [5]. The wide-angle field of view and variable shooting angles of UAV-acquired images enable coverage of complex terrains and environments. However, they also introduce challenges such as small target sizes, complex and diverse backgrounds, and increased detection difficulty. These factors result in inadequate feature map representation and hinder accurate extraction of critical information, thereby adversely affecting detection performance. As a result, enhancing the precision and reliability of detecting insulator defects in drone-captured images has emerged as a central topic of contemporary studies.

Deep learning methods have recently gained widespread attention in various engineering applications, demonstrating their effectiveness in tasks such as real-time electrical load prediction and management [6]. Interest in object detection has increased markedly, with algorithms generally classified as either one-stage or two-stage detectors [7]. In two-stage detection, the system initially selects regions that might contain objects and then inspects them thoroughly to classify the objects and refine the predicted bounding boxes. Prominent examples of this methodology, based on convolutional neural networks, include the Region-based Convolutional Neural Network (R-CNN) series of models, which have demonstrated high accuracy in both object localization and recognition tasks [8-11]. Despite their high accuracy, two-stage frameworks suffer from computational redundancy, which constrains their real-time applicability due to reduced inference efficiency.

One-stage detection algorithms perform target localization and classification directly and simultaneously on input images. The YOLO series and SSD are commonly regarded as representative architectures of one-stage object detectors [12-14]. Recent research has increasingly adopted YOLO-derived architectures to address insulator defect detection, yielding encouraging outcomes. For example, authors in [15] introduced MTI-YOLO, an enhanced version of the YOLO-tiny model specifically optimized for analyzing complex aerial images of insulators. Their method incorporates multi-scale feature aggregation and a SPP module to improve detection accuracy and robustness. Authors in [16] developed Bidirection-Fusion-YOLOv3, an improved YOLOv3-based method that utilizes a bidirectional fusion structure to strengthen feature representation and improve detection accuracy in challenging UAV-captured environments. Authors in [17] introduced an enhanced YOLOv5-based algorithm for insulator breakage detection, which incorporates ECA-Net attention within the backbone, BiFPN in the neck, and Soft-NMS in the prediction layer, achieving a mAP of 95.02% and 49.4 FPS on aerial datasets, thereby satisfying real-time engineering requirements. Authors in [18] presented an improved YOLOv7-based method integrating K-means++ anchor clustering, Coordinate Attention, and HorBlock modules, alongside SIoU and focal loss functions to facilitate convergence and address sample imbalance in complex transmission line images. Authors in [19] proposed an enhanced YOLOv8s-based model featuring a MLKA module and a lightweight GSC_C2f structure based on GSConv, aimed at boosting multi-scale feature representation while reducing computational costs for UAV insulator defect detection. Authors in [20] developed a YOLOv9-based model targeting insulation breakage detection in column-mounted transformers using UAV imagery; this model integrates Haar wavelet downsampling, the iEMA attention mechanism, and a modified MPDIoU loss to improve detection accuracy. Authors in [21] introduced Bi-YOLOv10, an object detection model optimized for power inspection, which addresses sample imbalance and multiscale complexity in UAV images by incorporating a BiFPN module for efficient feature fusion and a Focal-EIoU loss function to balance contributions from high- and low-quality samples. Authors in [22] developed an enhanced algorithm based on YOLOv11 to detect insulator defects in power distribution systems. Their model incorporates ASFF for effective multi-scale feature fusion, utilizes a BiFPN structure augmented with CBAM attention to strengthen contextual feature learning, and adopts ShuffleNetV2 as a lightweight backbone to reduce computational cost and enable efficient deployment. Authors in [23] introduced FPFS-YOLO, an enhanced insulator defect detection framework built upon the YOLO11n architecture. The model leverages the lightweight FasterNet as its backbone and integrates the ParNet attention mechanism to improve recognition of small-scale defects. It combines a PSP-Head structure with a specific detection head for tiny targets (C3k2_faster module) to further optimize performance, striking a compromise between computational efficiency and detection accuracy.

In practical UAV-based inspections of power transmission lines, high-resolution cameras are typically used to capture images from a considerable distance, resulting in a relatively small pixel proportion occupied by insulator targets. Moreover, variations in shooting angles and flight altitudes further exacerbate the inconsistency in target scales, posing significant challenges for standard object detection algorithms in terms of small object recognition, feature extraction, and scale adaptability. Additionally, the interference caused by background clutter—such as trees, the sky, and transmission infrastructure—poses significant challenges to small-target detection, increasing the likelihood of misclassification and omission. Conversely, conventional bounding box regression losses like Complete IoU (CIoU) and Generalized IoU (GIoU) often suffer from positional instability and sensitivity to spatial deviations, particularly in scenarios involving small objects or targets with extreme aspect ratios. Such limitations hinder further gains in detection precision.

Aiming to mitigate the aforementioned shortcomings, the proposed model enhances the baseline YOLOv11n by embedding the Conv2Former attention mechanism and

introducing PW-IoU, a novel loss function tailored for precise regression tasks. By fusing convolutional neural networks' local feature modeling with Transformers' global attention mechanisms, Conv2Former significantly improves the network's understanding and semantic encoding of small targets within cluttered backgrounds. Meanwhile, the PW-IoU loss function incorporates the prediction deviation modeling from PIoUv2 and the dynamic gradient reweighting mechanism from WIoUv3, which collectively improve bounding box regression precision and enhance the model's adaptability to variations in object position, shape, and scale. The collaboration between these two modules significantly improves the model's capability in detecting small objects and enhances its overall robustness. Importantly, this performance gain is achieved with minimal increase in computational cost, making the approach well-suited for reliable and efficient UAV-based inspection of power transmission infrastructure.

II. METHODOLOGY

A. Architecture of C2F-YOLOv11n Units

YOLOv11n in this work is derived from the Ultralytics YOLO framework (v8.3.59) and instantiated using a user-defined YAML configuration (yolov11-Conv2Formers.yaml) to ensure reproducibility. YOLO11n, an evolved version within the YOLO framework, introduces architectural refinements to both its backbone and neck through the integration of key components including CBS, C3k2, SPPF, and C2PSA. These enhancements notably improve feature extraction for small targets and scenes with high visual complexity. The model takes 640 × 640 input images and employs mosaic data augmentation to effectively improve robustness and adaptability to diverse scenarios. The neck network achieves multi-scale feature fusion through cross-layer connections, whereas the detection head utilizes depthwise separable convolutions to reduce parameter count and computational cost, simultaneously enhancing generalization ability and detection stability. With an excellent balance between accuracy and inference speed, YOLO11n is selected as the baseline model in this study to fully leverage its efficient performance and application potential in real-time object detection tasks.

YOLOv11n by default uses the C3k2 module, which effectively captures local spatial features through multiple convolutional layers and residual connections. However, the fixed kernel sizes within C3k2 limit the model's ability to perceive large-scale contextual information, leading to insufficient capacity in modeling global dependencies and long-range spatial relationships. Additionally, as the network deepens, convolutional modules in C3k2 may encounter gradient vanishing and performance degradation during training, which adversely affects generalization ability and convergence speed.

Since 2020, Vision Transformers (ViTs) have achieved remarkable results in image classification tasks. Despite their strong capability in capturing global dependencies through self-attention, they suffer from high computational overhead when dealing with high-resolution images. To overcome this limitation, researchers have proposed hybrid approaches that integrate convolutional operations with Transformer architectures. Among them, Conv2Former stands out as an effective convolutional Transformer model that replaces traditional self-attention with convolutional modulation [24]. This design significantly lowers computational costs while retaining the benefits of global information modeling, thereby enhancing the efficiency of high-resolution image processing, as illustrated in Figure 1.

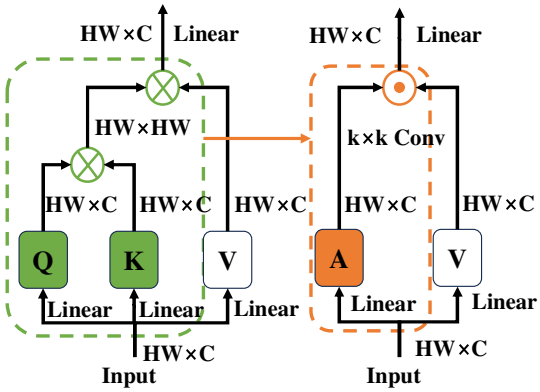

Fig. 1. Self-attention module with convolutional modulation.

As illustrated in Figure 2, Conv2Former is a Transformer–convolution hybrid network featuring a pyramidal architecture. The architecture is organized into four progressive stages, each featuring a unique configuration of convolutional blocks and specific feature map resolutions. Stage transitions are facilitated through Patch Embedding modules, which downsample the spatial dimensions while preserving essential semantic information. This architecture integrates convolutional operations for local feature extraction with Transformer modules for capturing global semantic information. By dynamically adjusting the weighting of convolutional feature responses, the self-attention mechanism effectively strengthens the encoding of features across multiple scales and orientations.

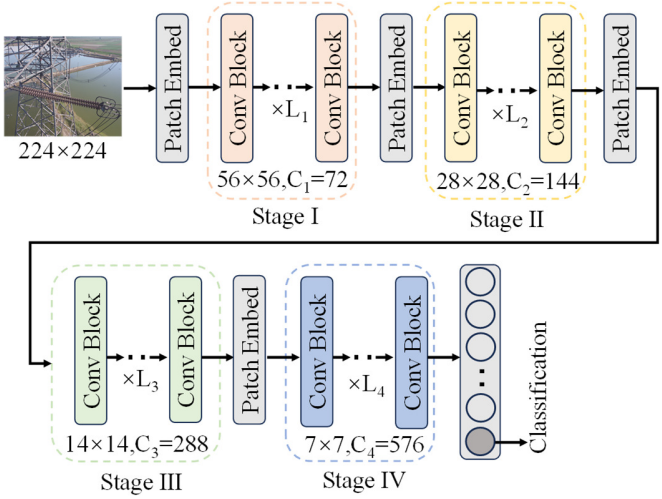

Fig. 2. Architecture of the Conv2Former.

Taking advantage of these improvements, four C3K2 modules in the YOLOv11n backbone were replaced with Conv2Former blocks. This modification led to the creation of the C2F-YOLOv11n architecture, as illustrated in Figure 3. With this enhancement, the network gains a heightened ability to capture global context and semantic cues. It also achieves a more balanced trade-off between computational cost and model complexity, thereby improving the accuracy of small object detection.

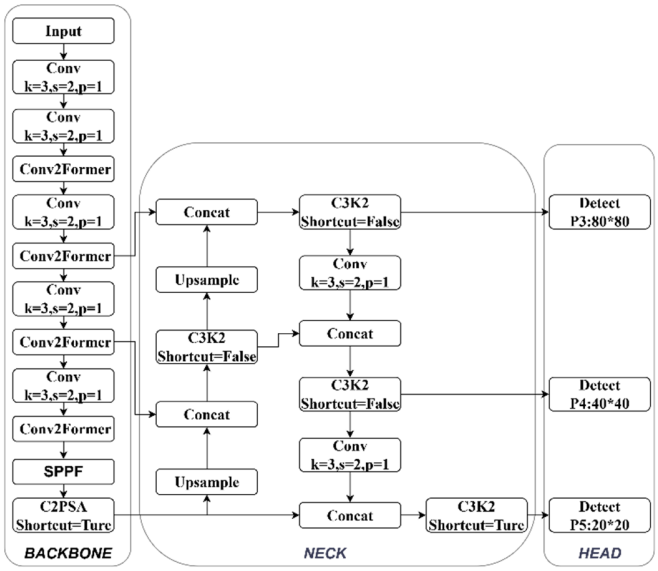

Fig. 3. Overall architecture of the C2F-YOLOv11n.

B. PW-IoU Loss Function

In object detection, the design of the bounding box regression loss plays a pivotal role in guiding the network to accurately locate objects. This loss function is designed to guide the predicted bounding boxes toward their corresponding ground truth annotations, thereby improving the precision and reliability of spatial localization.

Metrics like IoU serve as standard loss functions to quantify the similarity between predicted bounding boxes and their corresponding annotations, reflecting the accuracy of object localization. However, they often suffer from optimization instability when the boxes do not intersect. To overcome these limitations, several enhanced IoU-based loss functions have been developed, including Efficient IoU (EIoU), Distance IoU (DIoU), and CIoU. These approaches integrate additional geometric constraints, such as center point distance and aspect ratio alignment, thereby facilitating faster convergence and yielding higher localization accuracy.

In this section, we propose a novel regression loss function, PW-IoU, which integrates the attention-guided anchoring capability of PIoUv2 with the dynamic gradient modulation mechanism of WIoUv3 [25, 26]. This design aims to enhance the robustness and convergence efficiency of the object detector, particularly in challenging scenarios such as UAV-based power line inspection, where small and low-quality targets under complex backgrounds are common. While PIoUv2 emphasizes reducing the boundary discrepancy between predicted and ground-truth boxes through a non-monotonic attention mechanism, WIoUv3 dynamically modulates gradient responses based on anchor quality, thereby alleviating the optimization bias caused by sample imbalance. By fusing these two mechanisms, our proposed PW-IoU loss inherits both precise localization and dynamic regularization.

To operationalize this integrated design, we formulate the PW-IoU loss function in a multi-step manner. Each element of the suggested approach is described in depth in the following steps, starting with a penalty term that corrects for the scale-adaptive edge misalignment between the ground-truth and forecast boxes.

- Step 1: This work presents a penalty term P that dynamically adjusts to the target size in order to overcome the issue of anchor box enlargement and the regression performance limits of current IoU-based loss functions:

$$P = \frac{1}{4}\left(\frac{d_{w1}}{w_{gt}} + \frac{d_{w2}}{w_{gt}} + \frac{d_{h1}}{h_{gt}} + \frac{d_{h2}}{h_{gt}}\right) \quad (1)$$

As seen in Figure 4, w_{gt} and h_{gt} specify the ground truth box's width and height, whereas d_{w1}, d_{w2}, d_{h1}, d_{h2} measure how far the corresponding edges of the predicted box deviate from the ground truth box.

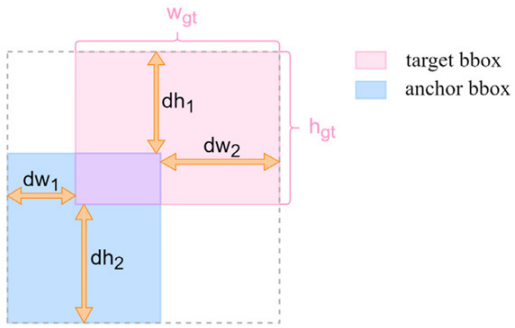

Fig. 4. Illustration of the penalty term P.

- Step 2: We apply an exponential decay function to the penalty factor P to obtain a weight q bounded between 0 and 1:

$$q = e^{-P}, \quad q \in (0,1] \quad (2)$$

Smaller boundary deviations yield values of q closer to 1, emphasizing well-localized predictions, whereas larger deviations reduce the weight, diminishing the impact of poor predictions during training.

- Step 3: To further refine gradient updates, we define a nonlinear modulation function:

$$u(\lambda q) = 3 * (\lambda q) * e^{-(\lambda q)^2} \quad (3)$$

where λ is a scaling hyperparameter that regulates the modulation sensitivity. This function balances gradient contributions by amplifying moderate weights and suppressing extreme values, thereby mitigating the imbalance between hard and easy samples.

- Step 4: Considering the quality of anchor boxes, we introduce a response function $R(Q)$ based on the IoU metric:

$$R(Q) = 1 + \beta * \log(Q + \varepsilon), \ Q = \text{detach}(\text{IoU}) \quad (4)$$

where Q is the detached IoU score preventing gradient backpropagation, β is a scaling coefficient, and ε is a small constant to ensure numerical stability. This function dynamically adjusts the gradient magnitude according to the anchor box quality, promoting more effective learning from higher-quality anchors.

- Step 5: Integrating the above components, the proposed PW-IoU loss function is formulated as:

$$L(Q) = R(Q) * u(\lambda q) * \left(L_{\text{IoU}} + 1 - e^{-P^2}\right) \quad (5)$$

where L_{IoU} is the base IoU loss measuring the overlap between predicted and ground truth boxes. This composite loss function synergistically combines precise localization, adaptive gradient reweighting, and boundary deviation penalties, leading to improved convergence speed and detection accuracy in object localization tasks.

By combining the distance-sensitive focusing capability of PloUv2 and the quality-aware gradient adjustment of WIoUv3, the proposed PW-IoU loss improves the regression accuracy and convergence speed of the detector. Compared to existing IoU-based losses, our method demonstrates superior performance in insulator defect detection tasks.

C. Insulator Dataset and Experimental Details

The self-built dataset consists of 2,096 insulator images, including 976 images from the Insulator Defect Image Dataset (IDID) [27]. Additionally, 1,120 images were acquired on-site using UAVs during inspections of the power grid in the vicinity of Changsha, Hunan Province, China, between January and August 2024. The images encompass insulators under diverse conditions, including normal, flashover, and fracture, providing a comprehensive foundation for defect detection and classification tasks. Within the dataset, multiple instances are annotated, specifically containing 2,790 broken samples, 5,542 insulator samples, and 5,750 flashover samples. The dataset was organized into training, validation, and test sets, following a 7:2:1 division.

To guarantee experimental consistency and replicability, all tests were carried out using the same hardware and software setup. The proposed model was trained and evaluated in a cloud computing environment with GPUs and ample memory, ensuring efficient computational performance. The model training utilized PyTorch version 2.1.0 with Python 3.11. The experiments were run on an Ubuntu system, powered by a Xeon(R) Silver 4214R CPU and a NVIDIA RTX 3080 GPU. Additional training parameters are detailed in Table I.

TABLE I. EXPERIMENTAL DETAILS

Parameter	Value	Parameter	Value
Batch size	8	Optimizer	SGD
Initial learning rate	0.01	Weight decay	0.0005
Final learning rate	0.01	Epoch	200

D. Performance Evaluation Metrics

To comprehensively evaluate performance, this study adopts six key metrics: Precision, Recall, F1-score, mAP50, mAP50-95, and FPS. These metrics measure the model's overall performance in important areas, including generalization ability and detection reliability [28].

The percentage of true positive samples among all samples that were anticipated to be positive is known as Precision:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (6)$$

In (6), TP corresponds to the quantity of accurately detected positive samples, whereas FP signifies the quantity of falsely identified positive samples.

Recall represents the fraction of actual positive instances that are successfully identified by the model. It indicates how thoroughly the model captures all relevant positive cases, reflecting the coverage of its detection capability. The recall metric is computed as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (7)$$

In (7), the term TP corresponds to positive examples accurately classified by the model, whereas FN captures the positive instances that remained undetected.

The F1-score (F1), calculated as the harmonic average of Precision and Recall, provides a unified performance indicator that balances the strengths and limitations of both metrics. This measure is particularly valuable in scenarios with imbalanced class distributions, where it delivers a more reliable evaluation of the model's effectiveness in detecting underrepresented categories. The computation of F1 is expressed as follows:

$$\text{F1} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision+Recall}} \quad (8)$$

When there is a substantial imbalance between Precision and Recall, the F1 tends to fall significantly below their arithmetic mean. This discrepancy suggests that the model's decision boundaries are suboptimal and require further refinement.

AP measures the overall detection performance of a model for a single class, capturing its ability to maintain precision across varying recall thresholds. It is computed by integrating the area under the Precision–Recall (P–R) curve, thus reflecting how well the model balances Precision as Recall increases:

$$\text{AP} = \int_0^1 \text{P(r)}\, \text{dr} \quad (9)$$

In (9), P(r) denotes the precision at a given recall level r, representing the proportion of correctly identified positive samples among all detections at that recall threshold. The variable corresponds to the recall rate, varying continuously from 0 to 1, indicating the fraction of true positive samples successfully retrieved by the model.

mAP serves as a comprehensive performance metric that captures the model's precision across varying recall levels. The calculation involves integrating the P–R curve for each individual class to obtain the AP, followed by averaging these

AP values over all classes to yield the final mAP score. The precise computational formula is as follows:

$$\text{mAP} = \frac{\sum_{i=1}^{C} \text{AP}_i}{C} \quad (10)$$

In (10), AP_i represents the average precision associated with the i -th class, whereas C indicates the total number of categories included in the evaluation process.

mAP50 and mAP50-95 are important indicators for assessing object detection models. The mAP50 metric evaluates detection performance using a relaxed IoU threshold of 0.50, highlighting the model's ability to identify objects. Meanwhile, mAP50-95 averages precision scores across multiple IoU thresholds from 0.50 to 0.95 in 0.05 increments, offering a more rigorous measure of the model's accuracy in object localization and its overall robustness across different overlap requirements.

FPS refers to the rate at which an imaging or video system handles individual frames each second, serving as a key indicator of processing performance.

$$\text{speed} = \frac{T_{test}}{\text{num}} \quad (11)$$

$$\text{FPS} = \frac{1}{\text{speed}} \quad (12)$$

where T_{test} denotes the total inference time required by the model to process the complete test dataset, whereas num represents the total number of images contained within the dataset.

III. RESULTS AND DISCUSSION

A. *Ablation Experiments*

Comparative tests were carried out using the default CIoU loss function against the baseline YOLOv11n under the same conditions in order to verify the efficacy of the suggested C2F-YOLOv11n model. The findings, which are compiled in Table II, show the superiority of the suggested model in object detection tasks through a quantitative assessment across several important performance indicators.

TABLE II. ABLATION EXPERIMENTS FOR INSULATOR DEFECT DETECTION

Model	YOLOv11n	C2F-YOLOv11n
Precision (%)	90.3	91.7
Recall (%)	82.6	83.1
F1 (%)	86.3	87.2
mAP50 (%)	88.2	89.4
mAP50-95 (%)	58	58.9
FPS	176	194
Model size (MB)	2.59	2.7

We conducted ablation experiments under identical training conditions with CIoU loss to comprehensively evaluate the performance of the proposed C2F-YOLOv11n. As reported in Table II, incorporating Conv2Former into the YOLOv11n backbone consistently enhanced all evaluation metrics. In particular, Precision increased by 1.4%, Recall by 0.5%, and F1-score by 0.9%. Moreover, mAP50 and mAP50-95 rose by 1.2% and 0.9%, demonstrating improved accuracy in detecting small and visually ambiguous objects. These gains can be attributed to the design of the Conv2Former module, which combines convolutional operations with lightweight Transformer-inspired attention. In C2F-YOLOv11n, four C3k2 modules in the backbone were replaced with Conv2Former blocks, strengthening feature representation and improving inference efficiency. This modification led to an 18 FPS increase over the baseline, primarily due to Conv2Former's efficient convolutional modulation, which reduces redundant computations while preserving global contextual information. Although the model size increased slightly by 0.11 MB, the trade-off is justified by the significant gains in accuracy and robustness, making C2F-YOLOv11n highly suitable for real-time object detection in complex scenarios such as UAV-based insulator defect inspection.

Figure 5 illustrates the outcomes of the ablation studies, showcasing the Recall, Precision, mAP50, and mAP50-95 curves. The data clearly indicate that the C2F-YOLOv11n model achieves quicker convergence throughout training and consistently surpasses the baseline model in performance. These findings strongly confirm the proposed C2F-YOLOv11n model's effectiveness and its superiority in detecting insulator defects.

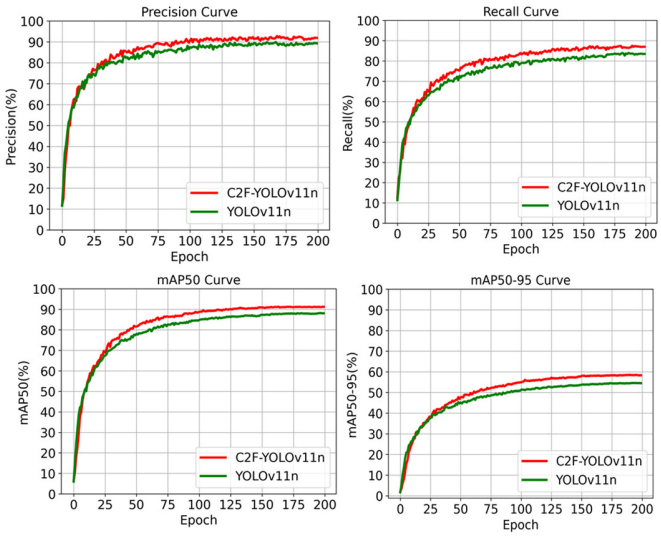

Fig. 5. Curves of Recall, Precision, mAP50, and mAP50-95 for ablation experiments.

To directly assess the detection capabilities of the enhanced C2F-YOLOv11n model, four images were randomly selected from the dataset for inference testing. As illustrated in Figure 6, the first row shows the detection results of the original YOLOv11n model, whereas the second row presents those of the improved C2F-YOLOv11n model. The enhanced model significantly decreased false positives and missed detections by correctly identifying all defect targets. Compared to the original YOLOv11n, the C2F-YOLOv11n demonstrates superior detection accuracy and overall performance, particularly demonstrating stronger recognition ability when detecting small targets.

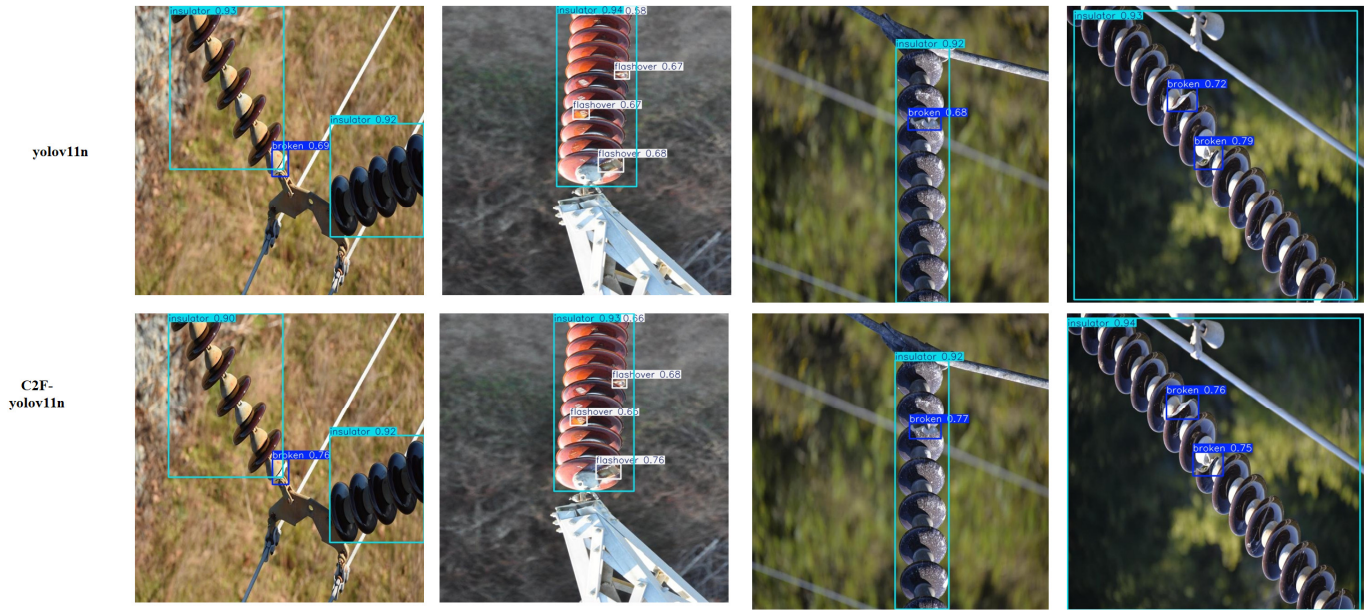

Fig. 6. Visual comparison of YOLOv11n and C2F-YOLOv11n on the self-built insulator dataset.

B. Comparative Experiments

A number of comparative tests were carried out using the identical experimental setups and dataset conditions in order to confirm the benefits of the suggested C2F-YOLOv11n model in object detection tasks. The model was evaluated against mainstream YOLO-based detection algorithms, with the results summarized in Table III.

TABLE III. COMPARISON OF PERFORMANCE METRICS FOR DIFFERENT YOLO-BASED ALGORITHMS

Model	Precision (%)	Recall (%)	mAP50-95 (%)	FPS	Model size (MB)
YOLOv8n	88.6	80.7	56.7	137	3.01
YOLOv10n	87.0	81.3	57.5	165	2.71
C2F-YOLOv11n	91.7	83.1	58.9	194	2.70

From the experimental results in Table III, it can be observed that although YOLOv8n and YOLOv10n exhibit relatively fast inference speeds of 137 FPS and 165 FPS, respectively, their detection accuracy remains suboptimal, with mAP50-95 scores of only 56.7% and 57.5%. In contrast, the proposed C2F-YOLOv11n model achieves the highest Precision (91.7%), Recall (83.1%), and mAP50-95 (58.9%) among the compared models, while maintaining a competitive inference speed of 194 FPS and a compact model size of 2.70 MB.

Compared to other lightweight YOLO variants that prioritize speed or model compression, the C2F-YOLOv11n model demonstrates a more balanced performance in terms of both accuracy and efficiency. This highlights its suitability for real-time UAV-based insulator defect detection tasks, where both precision and responsiveness are crucial.

C. Loss Function Comparison Test

In order to assess how regression loss functions influence network performance, this work conducts a comparative analysis of typical IoU-based approaches—DIoU, EIoU, Powerful-IoU, and Wise-IoU—against the baseline CIoU applied in C2F-YOLOv11n, together with the newly introduced PW-IoU. Using C2F-YOLOv11n as the reference framework, the experimental outcomes are reported in Table IV, with "Y+xIoU" representing the baseline model enhanced by the respective IoU-based loss functions.

TABLE IV. COMPARISON OF IOU-BASED REGRESSION LOSS FUNCTIONS

Model	Precision (%)	Recall (%)	mAP50 (%)	mAP50-95 (%)
Y+ CIoU	91.7	83.1	89.4	58.9
Y+ DIoU	92.5	86.2	90.8	57.3
Y+ EIoU	92.3	85.8	90.7	56.9
Y+ Powerful-IoU	90.1	84.9	90.3	60.2
Y+ Wise-IoU	93.2	85.5	90.9	59.9
Y+ PW-IoU	92.6	87.5	90.8	59.3

As shown in Table IV, the baseline model with CIoU achieved a mAP50 of 89.4% and a mAP50-95 of 58.9%. Integrating Wise-IoU and Powerful-IoU led to moderate improvements in mAP50 (90.9% and 90.3%, respectively), although their mAP50-95 remained relatively low (59.9% and 60.2%). In contrast, the model employing the proposed PW-IoU achieved the highest Recall of 87.5%, representing a 4.4 percentage point improvement over CIoU. Although PW-IoU's mAP50-95 (59.3%) is marginally lower than that of Powerful-IoU and Wise-IoU, it provides superior recall performance, which is crucial for the detection of small or partially occluded insulator defects. This demonstrates a trade-off between maximizing Recall for comprehensive defect detection and achieving high localization accuracy, where other IoU-based losses slightly outperform. Therefore, PW-IoU was selected for

its balanced performance in both robust detection and reliable bounding box regression. This choice demonstrates the method's applicability for small-object and complex-background scenarios while ensuring high reliability of detection, providing an effective solution for UAV-based insulator defect inspection.

Figure 7 presents a visual comparison of detection outcomes using the C2F-YOLOv11n model trained with various IoU-based regression loss functions, including CIoU, DIoU, EIoU, Powerful-IoU, Wise-IoU, and the proposed PW-IoU. As shown, all models are capable of accurately detecting the main insulator body. However, significant disparities are observed in detecting small-scale damaged areas (labeled as "broken"). Specifically, models utilizing CIoU, DIoU, and EIoU sometimes produce lower confidence scores or miss detections altogether. While Powerful-IoU and Wise-IoU show improved consistency, they still fall short in precise localization. In contrast, the model trained with the proposed PW-IoU demonstrates superior performance, achieving consistently high confidence and accurate bounding boxes across all defect instances. This highlights the enhanced small-object detection and regression capabilities of PW-IoU, confirming its effectiveness and robustness in real-world power equipment inspection tasks.

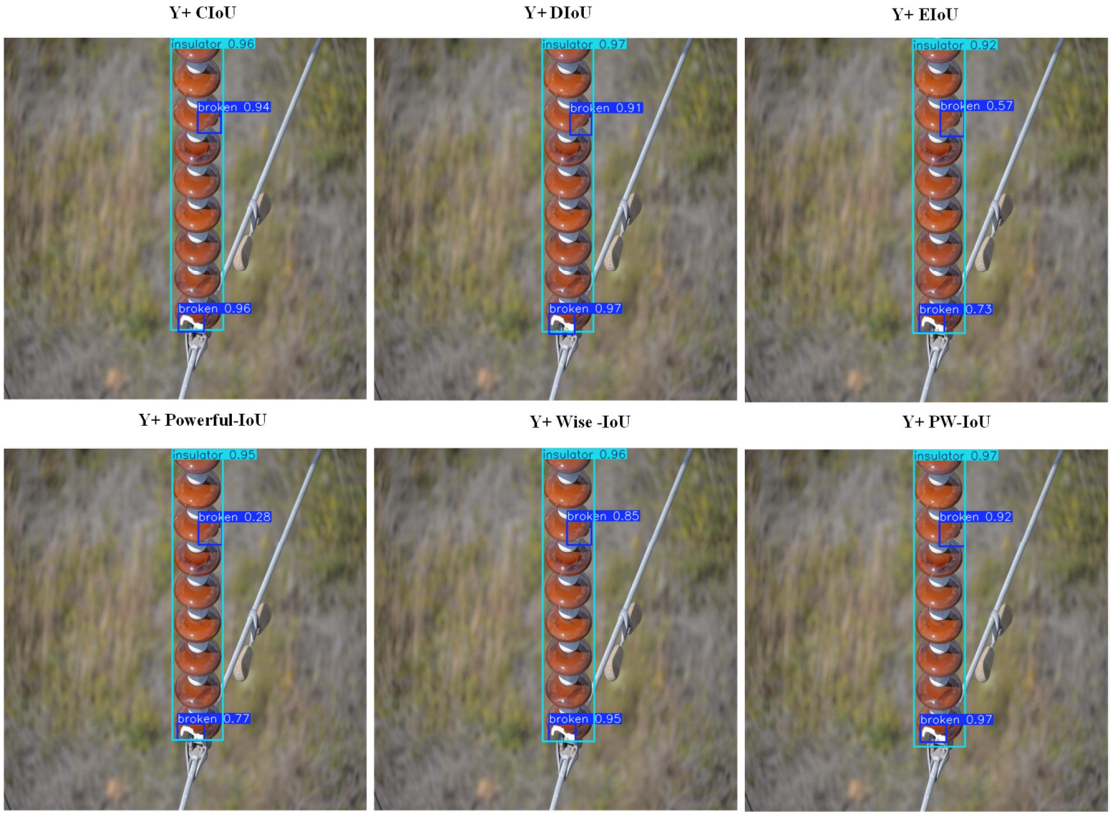

Fig. 7. Visual comparison of detection results using different IoU-based regression loss functions.

These findings show that in the task of insulator defect identification, PW-IoU performs noticeably better than the original CIoU and other state-of-the-art IoU-based loss functions. The performance improvement is primarily attributed to the fusion of PIoUv2's boundary-aware localization and WIoUv3's gradient-aware optimization. The constraints of conventional IoU-based loss functions in intricate UAV-captured inspection settings are successfully overcome by PW-IoU, which improves the model's capacity to identify tiny or hazy insulator flaws by dynamically weighting samples based on boundary deviation and anchor quality.

D. Experimentation on the Chinese Power Line Insulator Dataset

To assess the generalization capability of the proposed model, comparative experiments were conducted on the publicly available Chinese Power Line Insulator Dataset (CPLID) for insulator defect detection [29]. The dataset contains 848 images in two categories: normal and defective insulators. The detection performance of C2F-YOLOv11n combined with the PW-IoU loss function was evaluated, and the detailed results are presented in Table V.

TABLE V. COMPARISON OF RESULTS ON THE CPLID DATASET

Model	Precision (%)	Recall (%)	F1 (%)	mAP50 (%)
Approach in [30]	96.9	94.3	95.6	97.5
Approach in [31]	95.3	95.7	95.5	98.7
Approach in [32]	95.4	85.8	90.3	91.2
Approach in [33]	95.4	86.0	90.5	90.9
C2F-YOLOv11n	96.5	95.5	96.0	97.9

As shown in Table V, C2F-YOLOv11n in conjunction with the PW-IoU loss function achieves an F1-score of 96.0% and a mAP50 of 97.9% on the CPLID dataset. These results indicate not only robust detection performance but also practical applicability for reliable insulator defect inspection in real-world scenarios.

E. Discussion

The experimental outcomes indicate that integrating Conv2Former into the YOLOv11n backbone significantly improves feature representation, particularly for small and low-contrast insulator defects in complex aerial backgrounds. This enhancement supports the model in capturing both local details and global contextual information, which is critical for robust detection in UAV imagery, where targets often appear small and partially occluded.

The PW-IoU loss function further strengthens the model's bounding box regression by dynamically adjusting the gradient based on prediction quality and boundary alignment. Although it does not achieve the highest mAP50-95 compared to some other IoU-based losses, it excels in Recall, which is crucial for ensuring that small or partially occluded defects are detected. This trade-off demonstrates that prioritizing Recall may be more beneficial in UAV-based insulator inspection tasks, where missed detections can have significant safety implications, whereas minor losses in localization precision are acceptable. This adaptive weighting mechanism effectively mitigates issues arising from sample imbalance and varying target scales, leading to improved convergence stability and localization precision. The consistent performance gains across multiple metrics underscore the synergy between architectural improvements and the novel loss design. Moreover, comparative experiments on the CPLID demonstrate that C2F-YOLOv11n with PW-IoU achieves competitive detection performance, further confirming its strong generalization capability.

However, despite the promising results, some challenges remain. The increased model complexity due to the Conv2Former module, although modest, may still impose limitations on deployment in highly resource-constrained embedded UAV platforms. Moreover, the model's performance on extremely small or heavily occluded defects could be further improved by incorporating temporal information from video sequences or multi-view fusion techniques. Future work should also explore the integration of domain adaptation strategies to enhance robustness against varying environmental conditions such as different lighting, weather, and geographic regions. Additionally, lightweight model compression and pruning methods could be investigated to optimize the trade-off between accuracy and computational overhead, ensuring efficient real-time deployment without sacrificing detection quality.

IV. CONCLUSION

An improved YOLOv11n model, C2F-YOLOv11n, is introduced for detecting insulator defects in Unmanned Aerial Vehicle (UAV) transmission line images with high precision. The model introduces two core innovations: (1) the integration of the Conv2Former module into the YOLOv11n backbone, effectively enhancing global context perception and semantic feature extraction for small and blurry targets; and (2) the design of a novel regression loss function, PW-IoU, which combines PIoUv2's boundary-aware focusing and WIoUv3's quality-sensitive gradient reweighting to improve localization accuracy and convergence stability.

Extensive evaluation shows that C2F-YOLOv11n outperforms existing lightweight YOLO variants, offering a favorable balance of accuracy, speed, and model compactness. These results highlight its suitability for real-time insulator defect detection in UAV-collected images. Furthermore, comparative experiments on different IoU-based loss functions show that PW-IoU achieves superior performance, with a Recall of 87.5%, outperforming Complete IoU (CIoU), Distance IoU (DIoU), Efficient IoU, Powerful-IoU, and Wise-IoU. The combination of Conv2Former and PW-IoU allows the model to better handle the challenges of small target detection, scale variation, and complex backgrounds common in UAV inspection scenarios. With high detection accuracy and real-time inference capability, the proposed model provides a robust and practical solution for intelligent power line monitoring. Moreover, experiments on the Chinese Power Line Insulator Dataset (CPLID) confirm the model's robustness and generalization across different insulator images, supporting its practical applicability in diverse UAV inspection scenarios.

Future efforts will extend the model to cover a broader spectrum of insulator defects and environmental conditions, aiming to improve its generalization performance. In addition, deployment on embedded UAV platforms will be explored to evaluate the real-world applicability of the proposed algorithm under resource-constrained conditions.

ACKNOWLEDGMENT

This research was financially supported by Mahasarakham University, Thailand.

REFERENCES

[1] E. Antwi-Bekoe, G. Liu, J.-P. Ainam, G. Sun, and X. Xie, "A deep learning approach for insulator instance segmentation and defect detection," *Neural Computing and Applications*, vol. 34, no. 9, pp. 7253–7269, May 2022, https://doi.org/10.1007/s00521-021-06792-z.

[2] Y. Ma, Q. Li, L. Chu, Y. Zhou, and C. Xu, "Real-Time Detection and Spatial Localization of Insulators for UAV Inspection Based on Binocular Stereo Vision," *Remote Sensing*, vol. 13, no. 2, Jan. 2021, Art. no. 230, https://doi.org/10.3390/rs13020230.

[3] D. Doufene, S. Benharat, S. Bouazabia, and S. A. Bessedik, "Hybrid Grey Wolf and Finite Element Method (GWO-FEM) Algorithm for Enhancing High Voltage Insulator String Performance in Wet Pollution Conditions," *Engineering, Technology & Applied Science Research*, vol. 12, no. 3, pp. 8765–8771, June 2022, https://doi.org/10.48084/etasr.4978.

[4] S. Deng, L. Chen, and Y. He, "Insulator defect detection from aerial images in adverse weather conditions," *Applied Intelligence*, vol. 55, no. 6, Jan. 2025, Art. no. 365, https://doi.org/10.1007/s10489-025-06280-0.

[5] Z. Deng, X. Li, and R. Yang, "RML-YOLO: An Insulator Defect Detection Method for UAV Aerial Images," *Applied Sciences*, vol. 15, no. 11, June 2025, Art. no. 6117, https://doi.org/10.3390/app15116117.

[6] S. A. Ahmed, E. H. Khalifa, A. F. A. Mahmoud, F. A. Abdalla, M. Nawaz, and A. Sulman, "Real Time Electrical Load Prediction and Management through Deep Learning and Reinforcement Learning Techniques," *Engineering, Technology & Applied Science Research*,

vol. 15, no. 2, pp. 21061–21067, Apr. 2025, https://doi.org/10.48084/etasr.9559.

[7] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object Detection With Deep Learning: A Review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019, https://doi.org/10.1109/TNNLS.2018.2876865.

[8] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving Into High Quality Object Detection," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 6154–6162, https://doi.org/10.1109/CVPR.2018.00644.

[9] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 142–158, Jan. 2016, https://doi.org/10.1109/TPAMI.2015.2437384.

[10] R. Girshick, "Fast R-CNN," in *2015 IEEE International Conference on Computer Vision*, Santiago, Chile, 2015, pp. 1440–1448, https://doi.org/10.1109/ICCV.2015.169.

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, June 2017, https://doi.org/10.1109/TPAMI.2016.2577031.

[12] Y. Liu, D. Liu, X. Huang, and C. Li, "Insulator defect detection with deep learning: A survey," *IET Generation, Transmission & Distribution*, vol. 17, no. 16, pp. 3541–3558, Aug. 2023, https://doi.org/10.1049/gtd2.12916.

[13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016, pp. 779–788, https://doi.org/10.1109/CVPR.2016.91.

[14] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector," in *14th European Conference on Computer Vision*, Amsterdam, Netherlands, 2016, pp. 21–37, https://doi.org/10.1007/978-3-319-46448-0_2.

[15] C. Liu, Y. Wu, J. Liu, and J. Han, "MTI-YOLO: A Light-Weight and Real-Time Deep Neural Network for Insulator Detection in Complex Aerial Images," *Energies*, vol. 14, no. 5, Mar. 2021, Art. no. 1426, https://doi.org/10.3390/en14051426.

[16] Z. Yang, Z. Xu, and Y. Wang, "Bidirection-Fusion-YOLOv3: An Improved Method for Insulator Defect Detection Using UAV Image," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, 2022, Art. no. 3521408, https://doi.org/10.1109/TIM.2022.3201499.

[17] G. Han, M. He, M. Gao, J. Yu, K. Liu, and L. Qin, "Insulator Breakage Detection Based on Improved YOLOv5," *Sustainability*, vol. 14, no. 10, May 2022, Art. no. 6066, https://doi.org/10.3390/su14106066.

[18] J. Zheng, H. Wu, H. Zhang, Z. Wang, and W. Xu, "Insulator-Defect Detection Algorithm Based on Improved YOLOv7," *Sensors*, vol. 22, no. 22, Nov. 2022, Art. no. 8801, https://doi.org/10.3390/s22228801.

[19] L. Zhang, B. Li, Y. Cui, Y. Lai, and J. Gao, "Research on improved YOLOv8 algorithm for insulator defect detection," *Journal of Real-Time Image Processing*, vol. 21, no. 1, Jan. 2024, Art. no. 22, https://doi.org/10.1007/s11554-023-01401-9.

[20] J. Han, S. Hao, S. Li, and J. He, "Improved YOLOv9-Based Insulation Breakage Detection of Column Transformer," in *2025 2nd International Conference on Smart Grid and Artificial Intelligence*, Changsha, China, 2025, pp. 844–848, https://doi.org/10.1109/SGAI64825.2025.11010027.

[21] M. Wu, G. Wang, W. Zhang, L. Wen, and H. Qu, "Bi-YOLOv10: Sample Imbalance-Aware Feature Fusion-Based Object Detection for Power Inspection," in *2024 IEEE First International Conference on Data Intelligence and Innovative Application*, Nanning, China, 2024, pp. 1–6, https://doi.org/10.1109/DIIA62678.2024.10871263.

[22] Y. Ji, D. Zhang, Y. He, J. Zhao, X. Duan, and T. Zhang, "Improved YOLO11 Algorithm for Insulator Defect Detection in Power Distribution Lines," *Electronics*, vol. 14, no. 6, Mar. 2025, Art. no. 1201, https://doi.org/10.3390/electronics14061201.

[23] Y. Chai, X. Yao, M. Chen, and S. Shan, "FPFS-YOLO: An Insulator Defect Detection Model Integrating FasterNet and an Attention Mechanism," *Sensors*, vol. 25, no. 13, July 2025, Art. no. 4165, https://doi.org/10.3390/s25134165.

[24] Q. Hou, C.-Z. Lu, M.-M. Cheng, and J. Feng, "Conv2Former: A Simple Transformer-Style ConvNet for Visual Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 12, pp. 8274–8283, Dec. 2024, https://doi.org/10.1109/TPAMI.2024.3401450.

[25] C. Liu, K. Wang, Q. Li, F. Zhao, K. Zhao, and H. Ma, "Powerful-IoU: More straightforward and faster bounding box regression loss with a nonmonotonic focusing mechanism," *Neural Networks*, vol. 170, pp. 276–284, Feb. 2024, https://doi.org/10.1016/j.neunet.2023.11.041.

[26] Z. Tong, Y. Chen, Z. Xu, and R. Yu, "Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism." arXiv, Apr. 08, 2023, https://doi.org/10.48550/arXiv.2301.10051.

[27] J. Liu, M. Hu, J. Dong, and X. Lu, "Summary of insulator defect detection based on deep learning," *Electric Power Systems Research*, vol. 224, Nov. 2023, Art. no. 109688, https://doi.org/10.1016/j.epsr.2023.109688.

[28] Y. Ye *et al.*, "TSSSKD-YOLO: an intelligent classification and defect detection method of insulators on transmission lines by fusing knowledge distillation in multiple scenarios," *Multimedia Systems*, vol. 31, no. 3, Apr. 2025, Art. no. 183, https://doi.org/10.1007/s00530-025-01772-y.

[29] X. Tao, D. Zhang, Z. Wang, X. Liu, H. Zhang, and D. Xu, "Detection of Power Line Insulator Defects Using Aerial Images Analyzed With Convolutional Neural Networks," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 4, pp. 1486–1498, Apr. 2020, https://doi.org/10.1109/TSMC.2018.2871750.

[30] T. Jiang, X. Hou, and M. Wang, "Insulator Defect Detection Based on the CDDCR–YOLOv8 Algorithm," *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, Sept. 2024, Art. no. 245, https://doi.org/10.1007/s44196-024-00654-x.

[31] L. Lu, L. Liu, J. Chen, J. Liu, and X. Wu, "YOLO-Lite: Lightweight Object Detection Algorithm Design and Optimization in Power System Inspection," *Signal, Image and Video Processing*, vol. 19, no. 9, June 2025, Art. no. 716, https://doi.org/10.1007/s11760-025-04312-3.

[32] Q. Zhang, J. Zhang, Y. Li, C. Zhu, and G. Wang, "IL-YOLO: An Efficient Detection Algorithm for Insulator Defects in Complex Backgrounds of Transmission Lines," *IEEE Access*, vol. 12, pp. 14532–14546, 2024, https://doi.org/10.1109/ACCESS.2024.3358205.

[33] Q. Zhang, J. Zhang, Y. Li, C. Zhu, and G. Wang, "ID-YOLO: A Multimodule Optimized Algorithm for Insulator Defect Detection in Power Transmission Lines," *IEEE Transactions on Instrumentation and Measurement*, vol. 74, 2025, Art. no. 3505611, https://doi.org/10.1109/TIM.2025.3527530.