Available online at www.ijournalse.org

Emerging Science Journal

(ISSN: 2610-9182)

Vol. 9, No. 6, December, 2025

YOLOv10-MsA: Attention-Augmented Real-Time Insulator Defect Detection from UAV Imagery

Junbiao Yang [1, 2], Nor Ashidi Mat Isa [1*]

[1] *School of Electrical and Electronic Engineering, Engineering Campus, Universiti Sains Malaysia, Pulau Pinang 14300, Malaysia.*

[2] *Hunan Institute of Traffic Engineering, Hengyang 421009, China.*

Abstract

The reliable operation of transmission systems depends on early detection of defects within high-voltage insulators because it helps stop power outages. Scalable inspection remains a challenge since UAV-based imagery and deep learning generate recent promising solutions. The goal of this study is to build an accurate and time-efficient insulator defect detection system through the YOLOv10 architecture integration of Manhattan Self-Attention (MsA), which strengthens spatial feature detection and increases robustness during evaluations in complex aerial inspection scenarios. The designers implemented the MsA modules into the backbone and neck sections of YOLOv10 to develop YOLOv10-MsA as their novel detection model. The model relied on 5,000 annotated insulator images acquired by unmanned aerial vehicles throughout various defect classes during training and evaluation. Standard object detection metrics consisting of mAP@0.5, mAP@0.5:0.95, precision, recall, F1-score, and inference speed evaluated the performance of the model. The YOLOv10-MsA system reached an mAP@0.5 performance of 93.1% and an F1-score of 91.9% at an inference speed of 79 FPS, which surpasses YOLOv8, YOLOv9, and baseline YOLOv10. The model demonstrated its best performance at detecting various small and hard-to-detect defects such as chipping and contamination. The application of MsA in detection systems resulted in better accuracy while preserving real-time operation, according to related model assessment. The proposed YOLOv10-MsA serves as a powerful deployable system for UAV-based insulator inspection because it achieves both high accuracy and fast operation. The method establishes conditions for real-time smart infrastructure observation with attention-augmented frameworks for object detection.

Keywords:

Insulator Defect Detection;

Yolov10;

Manhattan Self-Attention (MsA);

UAV Inspection;

Attention Mechanism;

Power Systems;

Real-Time Object Detection.

Article History:

Received:	06	May	2025
Revised:	27	October	2025
Accepted:	11	November	2025
Published:	01	December	2025

1- Introduction

Ensuring reliable electricity supply through power transmission systems requires a high standard of integrity to meet growing global energy demands. Within these systems, insulators serve as fundamental components, providing both mechanical support for conductors and electrical insulation between conductors and their high-voltage structures. Insulators are continuously exposed to harsh atmospheric and operational conditions, including ultraviolet radiation, industrial pollution, wind, dust, salt spray, and temperature fluctuations [1, 2]. These environmental stressors can cause physical deterioration, surface contamination, and ultimately lead to mechanical or dielectric failure. Defects such as cracks, chipping, flashovers, or pollution-induced corrosion compromise insulator reliability, reducing both operational safety and system performance. Such failures often trigger flashovers and arcing events, resulting in unexpected outages that pose risks to power infrastructure, as well as public and worker safety [3, 4]. Traditionally, insulator inspection has relied on manual visual examinations supplemented by infrared thermography. However, these methods are limited by human error and technical constraints [5].

* **CONTACT**: ashidi@usm.my

DOI: http://dx.doi.org/10.28991/ESJ-2025-09-06-02

In response, a technological shift has emerged, integrating unmanned aerial vehicles (UAVs) with computer vision and artificial intelligence to automate defect detection. High-resolution UAV imaging enables the capture of detailed aerial views of power lines and components [6]. When combined with deep learning algorithms, this approach offers a promising solution for rapid, accurate, and cost-efficient insulator monitoring [7].

Among deep learning models for object detection, the YOLO (You Only Look Once) family has emerged as the most prominent, offering simultaneous real-time detection and high accuracy. The evolution from YOLOv1 through YOLOv10 reflects continuous architectural advancements tailored to increasingly complex visual recognition tasks. YOLOv3 introduced the Darknet-53 backbone, while YOLOv4 incorporated Cross Stage Partial Network (CSPNet) and PANet for enhanced feature fusion. YOLOv5 further emphasized deployment [8, 9]. YOLOv6 and YOLOv7 focused on lightweight design and improved small-object detection, whereas YOLOv8 and YOLOv9 integrated anchor-free detection and transformer components for advanced spatial modeling [10, 11].

YOLOv10 introduced further innovations, including Non-Maximum Suppression (NMS)-free decoding, spatial–channel decoupled downsampling, large-kernel convolutions, and advanced label assignment strategies [12]. These improvements enhance performance in detecting both large and small objects while addressing challenges of sparse and imbalanced datasets. However, even with these advancements, YOLOv10 struggles with subtle or concealed defects in UAV imagery of insulators, as its convolutional structure relies on fixed receptive fields and struggles to capture long-range spatial dependencies [13]. To address this limitation, attention mechanisms such as the Convolutional Block Attention Module (CBAM) and Squeeze-and-Excitation Networks (SENet) have been integrated into CNN architectures, yielding improved feature extraction and localization performance [14, 15].

The Manhattan Self-Attention (MsA) mechanism has emerged as an efficient alternative to traditional dot-product attention. Unlike conventional approaches, MsA computes feature similarity using Manhattan distances, reducing computational overhead while retaining structural information. This design makes MsA particularly advantageous for complex aerial inspection environments [16].

This study investigates the integration of MsA into YOLOv10, producing the YOLOv10-MsA framework (Figure 1). By embedding MsA modules into both the backbone and neck structures, the model strengthens its ability to identify small, partially hidden, or visually obscured defects through multiscale attention mechanisms [17].

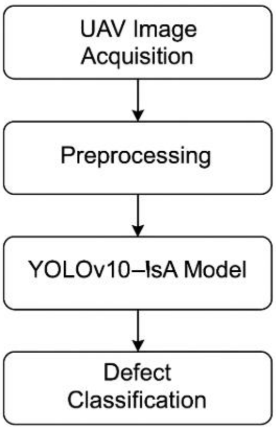

Figure 1. Overall Workflow of the Proposed System

The proposed model is evaluated using UAV-acquired datasets of insulator images annotated with various defect types, including flashovers, cracks, pollution, and structural failures. Performance is assessed using mean average precision (mAP), recall, precision, and frames per second (FPS). Experimental results demonstrate that YOLOv10-MsA achieves superior accuracy and faster inference speeds compared to YOLOv8, YOLOv9, and the baseline YOLOv10 [1].

In this article, the introduction section is followed by a brief description of research objectives and significance of this study, a literature review of section encompassing past articles that provided inspiration to this research, then followed by descriptions on the employed methodologies in this research. Results collected in this research is presented after the methodology section, tailed by an extensive discussion section, elaborating the impact of this research and concluded by revising the fulfillment of the objectives and hypotheses.

1-1- Research Objectives

The main purpose of this work involves developing a superior defect detection system for aerial imagery of insulators by employing Manhattan Self-Attention (MsA) within YOLOv10. The main research objective comprises the subsequent specific goals which follow:

- The objective is to create an efficient and accurate insulator defect detection system which merges YOLOv10 robustness with attention mechanisms spatial precision to support real-time transmission system fault detection.
- This research will examine the mechanism for adding Manhattan Self-Attention (MsA) into the YOLOv10 framework and its impact on visual representation features for tiny defects and objects that blend into outdoor scenes.
- A performance evaluation will take place for the proposed YOLOv10-MsA model by testing it against YOLOv8, YOLOv9, baseline YOLOv10 frameworks using mean average precision (mAP), precision, recall, and inference speed measurement standards.

This paper evaluates how attention-enhanced architectures affect infrastructure maintenance applications through real-time UAV-based aerial inspection tasks and analyzes deployment capabilities for such systems.

1-2- Significance of the Study

Automated infrastructure inspection techniques are enhanced by this research as it studies ways to detect insulator defects in high-voltage power systems to advance the current state-of-art solutions. This research holds multiple points of importance which become relevant throughout the study:

The automated insulator fault detection system through this approach helps preventive maintenance which reduces both outages and equipment failure and human injuries from manual inspections. YOLOv10 gains improved spatial perception by using Manhattan Self-Attention (MsA) at a reasonable computational expense which drives the development of efficient attention methods in computer vision.

A solution developed in this research enables accurate model deployment in real-time environments including UAV surveillance activities which monitor extended transmission line zones. The reduction in labor costs, equipment downtime, and emergency repair operations through early fault detection translates into significant savings for utility providers and infrastructure managers.

This work provides empirical evidence on the effectiveness of attention mechanisms in object detection and contributes to the literature on engineering applications of deep learning in aerial image analysis.

The remainder of this paper is organized as follows: Section 2 provides a literature review on object detection methods and recent developments in YOLO-based defect recognition. Section 3 describes the proposed methodology, including the integration of MsA into YOLOv10. Section 4 presents the dataset, training process, and results of experimental evaluations. Section 5 discusses the implications, limitations, and deployment strategies. Finally, Section 6 concludes with a summary and directions for future research.

2- Literature Review

The rapid evolution of object detection techniques over the last decade has had a profound impact on the automation of infrastructure monitoring, especially in the field of power transmission systems. Among the most critical elements in high-voltage electrical networks are insulators, which, due to their exposure to environmental stressors, are prone to degradation and failure [3, 7]. Traditional inspection methods, such as manual surveying or infrared thermography, are labor-intensive, costly, and subject to human error. The increasing availability of unmanned aerial vehicles (UAVs) and breakthroughs in computer vision have, however, opened new avenues for accurate, scalable, and automated insulator defect detection [18, 19].

2-1- Traditional and Early Machine Learning-Based Approaches

Early attempts at automating insulator monitoring primarily relied on handcrafted features in combination with classical machine learning classifiers. Methods such as histogram of oriented gradients (HOG), support vector machines (SVMs), and k-nearest neighbors (k-NN) were employed to categorize image patches based on surface texture, color distributions, and edge patterns [20, 21]. Although these approaches provided an initial foundation for automated inspection, they demonstrated limited robustness in real-world operating conditions, where the appearance of defects is influenced by fluctuations in lighting, background clutter, and variations in scale or orientation. Consequently, their applicability to large-scale aerial inspection tasks remained constrained.

The emergence of convolutional neural networks (CNNs) significantly addressed these shortcomings by enabling automated and hierarchical feature extraction. Unlike handcrafted methods, CNNs learn discriminative features directly from raw data, improving adaptability across diverse environments. Previous studies [22-24] confirm the transformative impact of CNN-based models across domains such as medical imaging, industrial manufacturing, and power system monitoring. Despite these advances, CNNs still encounter challenges when tasked with capturing long-range spatial dependencies and reliably detecting fine-scale objects from aerial imagery [25, 26]. These gaps highlight the necessity for further architectural innovations, particularly attention-based mechanisms, to enhance detection performance under complex environmental conditions.

2-2- Evolution of YOLO Architecture in Defect Detection

The field of real-time object detection has been most profoundly shaped by the You Only Look Once (YOLO) series of models, which consistently balanced speed and detection accuracy. Redmon & Farhadi [9] introduced YOLOv3, employing the Darknet-53 backbone to establish a strong foundation for single-shot detection. This model not only demonstrated the feasibility of combining accuracy with efficiency but also laid down architectural principles that later YOLO iterations adapted and refined.

Subsequent versions built upon this base by addressing key shortcomings. YOLOv4 and YOLOv5 integrated CSPNet with SPP modules to improve feature aggregation and surface detection [11, 27]. YOLOv6 shifted toward anchor-free detection, reducing computational load, while YOLOv7 enhanced multi-scale detection and small-object recognition through optimized detection heads [10, 12]. These advancements were crucial precursors to the authors' current focus on insulator defects, where small and partially obscured anomalies pose unique detection challenges.

The arrival of YOLOv8 introduced transformer-inspired backbones with anchor-free detection, showcasing stronger adaptability for complex datasets [28]. More recent iterations, YOLOv9 and YOLOv10 incorporated massive kernel convolutions, decoupled downsampling, and no-NMS decoding for more robust detection pipelines [29, 30]. While YOLOv10 delivers state-of-the-art performance across many domains, prior findings indicate it still struggles with fine-scale defect detection in UAV-based insulator imagery [31]. This gap underscores the motivation for the current study, where the integration of Manhattan Self-Attention (MsA) seeks to directly address the limitations of conventional YOLO backbones in localizing subtle, small-scale power system defects.

2-3- Attention Mechanisms in Object Detection

The development of attention mechanisms emerged as a direct response to the shortcomings of pure CNN-based structures, which, despite their success in feature extraction, often struggled to capture long-range dependencies and fine-scale details. Attention modules allow neural networks to selectively emphasize spatial or semantic features within an image, thereby improving the detection of objects that are either very small or partially occluded. Early contributions, such as the Convolutional Block Attention Module (CBAM) and Squeeze-and-Excitation Networks (SE-Net), demonstrated measurable gains in object localization and classification accuracy by embedding adaptive feature recalibration within CNN backbones [14, 15]. These studies highlighted that integrating attention into visual recognition pipelines could substantially enhance model sensitivity to subtle image patterns—an insight that directly informs the present authors' work on insulator defects.

Building on these foundations, researchers adapted transformer-based attention, originally developed for natural language processing into the visual domain, resulting in architectures capable of modeling global dependencies and complex contextual relationships [32, 33]. While highly effective, these approaches often incurred significant computational costs, making them impractical for deployment in real-time UAV applications. This limitation contrasts sharply with the requirements of insulator defect detection, where lightweight, fast, and resource-efficient models are essential.

The Manhattan Self-Attention (MsA) mechanism addresses this gap by replacing traditional dot-product similarity with Manhattan distance computations, thereby reducing overhead while retaining the benefits of localized feature interaction. As demonstrated in Feng et al. [16] and Heuillet et al. [34], MsA offers a balance between accuracy and efficiency, making it particularly well-suited for aerial inspection scenarios. The current study extends these insights by embedding MsA into the YOLOv10 architecture, positioning it as a more feasible solution for real-time defect detection in UAV-based power system monitoring.

2-4- YOLO and Attention Fusion in the Power Sector

A growing body of research has focused on integrating attention mechanisms into YOLO frameworks to address the persistent challenges of detecting small, subtle, or heavily occluded objects. For instance, Hu et al. [35] introduced DGW-YOLOv8, which combined deformable convolutions with the WIoU loss function to improve the detection of small-scale targets. Similarly, Ouyang et al. [6] proposed E-VarifocalNet, a lightweight architecture designed for power grid inspections, which leveraged adaptive attention mechanisms to identify difficult-to-spot insulator defects with notable efficiency. These studies collectively emphasized the potential of attention modules to improve detection accuracy in specialized scenarios.

Further refinements have also been reported. Li et al. [36] improved YOLOv8's bounding box regression for fault detection in insulators, thereby enhancing localization accuracy. In another line of work, Zhang et al. [1] applied a biformer-enhanced YOLOv8 to capture defects of varying sizes, confirming that hybrid attention designs could bolster recognition capability across diverse object scales. While these contributions underscore the importance of attention mechanisms in improving YOLO-based detectors, they also reveal a common limitation: many of the proposed solutions either increase computational complexity or lack robustness under real-world UAV constraints.

Against this backdrop, the current study positions the Manhattan Self-Attention (MsA) mechanism as a more resource-efficient alternative, offering the ability to process subtle spatial cues without substantially inflating model size or inference time. This efficiency makes MsA particularly relevant for real-time UAV-based insulator monitoring, distinguishing the present work from earlier studies that, while accurate, often sacrificed practicality in favor of model sophistication.

2-5- Dataset Availability and Annotation Challenges

The limited availability of comprehensive and well-annotated datasets remains a major barrier to the development of reliable insulator defect recognition systems. Current datasets suffer from two key shortcomings: many are confined to proprietary industrial repositories inaccessible to the broader research community, while others lack the scale and diversity necessary to train models capable of generalizing to complex real-world environments [37]. These restrictions hinder the development of robust algorithms that can adapt to the wide variability of UAV imagery encountered in field inspections.

Recent studies have attempted to address this gap by leveraging Open Images Dataset (OID) resources and UAV-acquired collections. However, annotation continues to pose a significant challenge, as accurate labeling often requires specialized domain expertise that is both time-consuming and costly [38, 39]. To overcome these challenges, researchers are increasingly turning to innovative annotation strategies. Two promising directions include the adoption of few-shot learning methods, which minimize the reliance on large annotated datasets by enabling models to generalize from limited samples [16], and the development of semi-supervised annotation frameworks, which combine a small amount of labeled data with abundant unlabeled samples to reduce manual annotation costs [40].

In the context of the present study, these limitations highlight the importance of designing detection architectures—such as the proposed YOLOv10-MsA—that are not only accurate but also capable of performing well in data-constrained settings, thereby extending their practical applicability to real-world UAV inspections.

2-6- Insulator-Specific Models and Loss Functions

Recent scholarship has increasingly focused on insulator fault detection using YOLO-based frameworks, reflecting the architecture's balance of speed and detection performance. For example, Guo et al. [20] explored the application of YOLOv3 to aerial imagery of insulators, but their findings highlighted persistent issues with background noise, which led to false positive detections and reduced reliability. Building upon this foundation, Liu et al. [5] introduced a lightweight YOLOv5s variant optimized for edge devices, demonstrating that compact models could be adapted for UAV-based inspections. This line of development was further advanced in [7, 41], who worked to improve both detection accuracy and inference speed, reinforcing YOLO's suitability for real-time monitoring applications.

Another key advancement has been the introduction of specialized loss functions to strengthen YOLO's performance. Techniques such as Distance-IoU (DIoU) and Complete-IoU (CIoU) improved the geometric interpretation of bounding box regression, thereby reducing localization errors during training [42, 43]. More recently, Zhang et al. [44] proposed the Focal Efficient IoU (EIoU) loss, which directs the learning process toward difficult cases, ensuring better robustness in challenging detection environments. These innovations have become integral to the latest YOLO variants, reflecting the field's ongoing effort to balance speed, precision, and robustness.

Within this trajectory, the current work situates YOLOv10-MsA as the next stage of evolution, combining advanced loss formulations with a novel Manhattan Self-Attention mechanism. This integration aims to not only address the persistent challenges of background noise and small-scale defect detection but also maintain operational efficiency in UAV-based inspections, a goal that earlier YOLO variations struggled to fully achieve.

2-7- Emerging Trends and Research Gaps

Recent advances in computer vision have highlighted the benefits of multi-branch architectures combined with spatial pyramid pooling and dynamic graph CNNs, which enable the extraction of richer and more discriminative features [2, 45]. Building on these innovations, researchers have introduced a new class of hybrid YOLO models that incorporate elements such as MobileViT backbones, integrated attention layers, and synthetic data augmentation strategies to improve robustness under resource-constrained environments [46-48]. These approaches underscore the field's ongoing push toward balancing lightweight design with enhanced representational power, particularly for UAV-based monitoring tasks.

Despite such progress, several critical challenges remain unresolved. In particular, real-time UAV imagery analysis continues to suffer from the computational burden of attention modules, with only limited attention-efficiency improvements reported in the literature. Existing attention-enhanced YOLO variants often fail to deliver the required level of field-ready performance, showing reduced inference speed and accuracy when applied outside of controlled experimental datasets (Table 1). Moreover, as emphasized in Zhang et al. [4] and Souza et al. [49] many proposed models are validated exclusively on idealized datasets, without adequate evaluation under diverse environmental conditions such as variable weather, occlusion, or complex backgrounds.

Table 1. Literature Review Matrix

Study	Model/Method	Contribution	Limitations
[9]	YOLOv3	Introduced fast, end-to-end object detection with Darknet-53	Poor performance on small objects
[18]	Deep Learning with CNN	Early use of CNN for insulator fault detection	Limited dataset; low accuracy in noisy environments
[42, 43]	DIoU, CIoU Loss Functions	Enhanced training convergence and bounding box accuracy	Computational cost is relatively high
[44]	Focal Efficient IoU Loss	Focused model training on hard examples	Requires careful parameter tuning
[6]	E-VarifocalNet	Lightweight model for UAV-based defect detection	Trade-off between speed and fine-grained accuracy
[5]	YOLOv5 Variants	Optimized for real-time UAV applications	Needs improved precision for small defect zones
[45]	CNN Architecture Review	Detailed overview of CNN optimizations	Lacks application to power infrastructure
[30]	YOLOv1–YOLOv8 + YOLO-NAS	Comprehensive review of YOLO architectures and use cases	No empirical testing on high-voltage assets
[48]	MobileViT-YOLO	Introduced lightweight attention-YOLO for insulators	Lower resolution models may miss fine-grained cracks
[3]	YOLO for Power Grid Monitoring	Applied YOLO to enhance grid reliability through automation	Model needs tuning for real-world conditions
[35]	DGW-YOLOv8 with Deformable Attention	Improved small object detection in insulators	Increased training complexity
[28]	Improved YOLOv8	Enhanced bounding box localization and precision	Limited generalization across environments
[1]	Biformer-YOLOv8	Boosted detection of multi-scale defects using attention modules	Increased memory footprint
[41]	Improved YOLOv8	Achieved higher mAP on custom insulator dataset	Dataset not diverse
[23]	CNNs in Industrial Defect Detectior	Reviewed CNN deployment in real-world applications	Not specific to aerial or insulator imagery
[31]	YOLO + UAV Monitoring	UAV-based real-time monitoring for transmission lines	Evaluation limited to a single region
[15]	CNNs in Vision	General review of CNNs in visual recognition	No application to UAV imagery or small objects
[38]	Segment Anything + DINO	Automated annotation with prompt learning	Early-stage approach, no validation on power datasets
[12]	Dynamic Graph CNNs with Spatial Pyramid	Achieved high accuracy on solar panel defect detection	Not tested in aerial imagery of insulators
[10]	YOLO Architecture Review	Surveyed YOLOv1 to YOLOv9 advancements	Lacks focus on embedded/real-time systems

These gaps directly motivate the current study's contribution. By embedding Manhattan Self-Attention (MsA) within YOLOv10, the authors address the dual challenge of maintaining computational efficiency while improving sensitivity to small-scale and obscured defects. Unlike prior hybrid YOLO frameworks, the proposed YOLOv10-MsA explicitly targets real-world UAV scenarios, thereby advancing the practical deployment potential of attention-based detectors in power system inspections.

3- Research Methodology

The research design part describes the development of an advanced model for high-voltage power insulator defect detection through YOLOv10 with Manhattan Self-Attention (MsA) integration. The methodology consists of architectural changes combined with dataset preparation and training setup and evaluation metrics because these elements work together to maximize real-time performance in UAV-based inspection tasks.

3-1- Base Architecture: YOLOv10

YOLOv10 serves as the core component in the proposed detection system because it maintains a good relationship between processing speed and detection accuracy. The newly introduced YOLOv10 within the YOLO family delivers multiple architectural upgrades that generate superior object localization performance when compared to YOLOv5 and YOLOv7 [29, 30]. The detection system contains four main advancements, which include dual label assignment for better ground truth matching, large-kernel convolution blocks for expansive receptive fields, spatial-channel decoupled downsampling for object boundary preservation, and non-maximum suppression-free decoding for latency reduction. YOLOv10 currently faces difficulties when it comes to detecting small-scale power line defects in high-resolution UAV images, which this research study addresses by integrating attention mechanisms.

3-2- Manhattan Self-Attention (MsA)

The study integrates the Manhattan Self-Attention (MsA) mechanism into the YOLOv10 backbone and neck to enhance spatial context feature extraction from the model. MsA provides speed and efficiency on lightweight operations because it calculates attention scores through L1 norm distance evaluation. The network structure enables focusing on important space areas while minimizing computational demands, which typically happen during attention operations. The exponential spatial distance decay of MsA weights enables the network to keep its attention on nearby and intermediate features, which help detect defects in complex backgrounds [16, 34]. Several shallow layers of this proposed architecture contain decomposed MsA modules to reduce memory requirements, while deeper layers are equipped with full MsA blocks to extract complex semantic information (Figure 2).

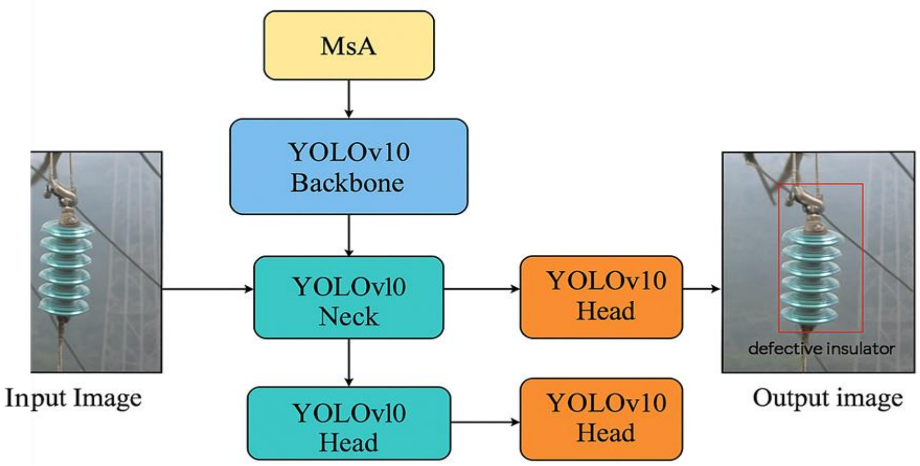

Figure 2. **Detailed architecture of YOLOv10 with Manhattan Self-Attention**

3-3- YOLOv10-MsA Model Architecture

YOLOv10-MsA embeds attention modules into the YOLOv10 architecture at two parts of the network during early and late stages. The backbone network contains two integration points for MsA modules at C2 and C3 stages to optimize spatial perception and object isolation for small and obscured defects. YOLOv10 maintains its dual-path prediction head structure for efficient training strategies to work properly (Figure 3). During training the model adopts Focal-EIoU loss as a function that focuses on difficult detection scenarios to enhance box regression performance [44].

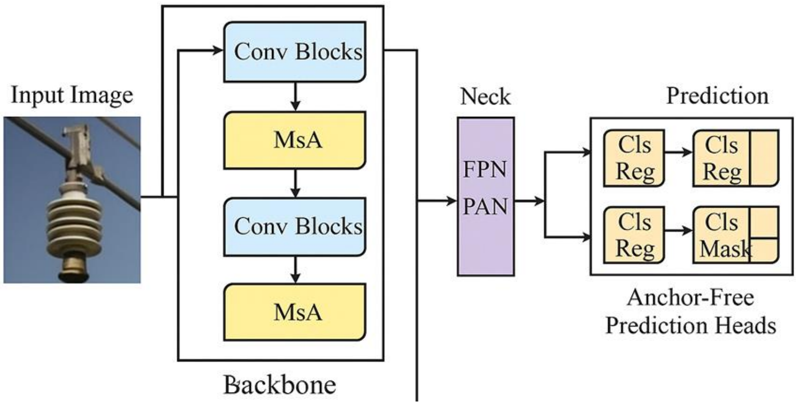

Figure 3. **Architecture of YOLOv10-MsA**

3-4- Dataset Description

For empirical validation, a custom dataset comprising aerial images of high-voltage transmission lines was compiled. The dataset consists of 5,000 labeled images captured by UAVs at various altitudes, angles, and lighting conditions. Each image is annotated using the YOLO bounding box format and includes both healthy and defective insulators categorized into four primary defect types: surface cracks, flashovers, contamination, and chipping. The dataset is partitioned into training (70%), validation (20%), and test (10%) subsets to ensure robust model evaluation. The experiment assessed model generalization by taking place in various environmental conditions, which included foggy conditions and both evening and dim lighting along with dense plant cover.

3-5- Training Protocol

The training process is executed using the PyTorch framework on an NVIDIA RTX 4090 GPU. Images are resized to 640×640 pixels to balance detection resolution and computational load. The Stochastic Gradient Descent (SGD) optimizer with a momentum value of 0.937 is used in conjunction with a one-cycle learning rate policy that peaks at 0.01. The model is trained for 200 epochs with a batch size of 16. Data augmentation techniques such as Mosaic, CutMix, random scaling, brightness jitter, and horizontal flipping are employed to simulate real-world variations and improve model robustness. Checkpoints are saved based on the highest validation mAP@0.5 to ensure optimal performance retention.

3-6- Evaluation Metrics

Model performance is assessed using a comprehensive set of metrics that reflect both accuracy and efficiency. Precision and recall are calculated to evaluate classification reliability, while the F1-score provides a harmonic mean of these two values. Mean Average Precision (mAP) is reported at IoU thresholds of 0.5 (mAP@0.5) and across a range from 0.5 to 0.95 (mAP@0.5:0.95) to measure detection quality. Inference speed, expressed in frames per second (FPS), is used to assess deployment feasibility in real-time UAV inspection applications. Model complexity, including parameter count and memory footprint, is also analyzed to verify suitability for edge devices.

3-7- Implementation Environment

All experiments are conducted in a controlled and reproducible environment using Ubuntu 20.04 LTS as the operating system. The software stack includes PyTorch v2.1.0 and CUDA v12.1. The MsA modules are implemented as modular PyTorch classes supporting variable kernel sizes (3, 5, and 7) and adaptable attention decay functions. Model training and validation scripts are configured for automatic check pointing, real-time metric logging, and visualization using TensorBoard. Final models are saved in ONNX format for cross-platform deployment on edge inference devices.

4- Results

This section presents the results of evaluating the proposed YOLOv10-MsA model for the task of insulator defect detection in UAV-captured aerial imagery. The performance of the model is compared against benchmark architectures, YOLOv8, YOLOv9, and baseline YOLOv10, using a custom dataset of high-voltage transmission lines. Evaluation metrics include mean average precision (mAP), precision, recall, F1-score, inference speed (FPS), and model complexity. Visual detection outputs and qualitative analyses are also provided to support the quantitative findings.

4-1- Performance Metrics

The proposed YOLOv10-MsA model demonstrates a clear advancement in detection performance when compared with prior state-of-the-art models. A detailed summary of the comparative results is presented in Table 2. At the Intersection over Union (IoU) threshold of 0.5, YOLOv10-MsA attained a mean average precision (mAP@0.5) of 93.1%, which substantially surpasses the baseline YOLOv10 (88.4%), as well as YOLOv9 (87.2%) and YOLOv8 (85.9%). The advantage of YOLOv10-MsA persists even under more stringent evaluation criteria. For instance, in terms of mAP@0.5:0.95, which requires robust performance across a range of IoU thresholds, YOLOv10-MsA achieved 72.6%, again outperforming YOLOv10 at 68.1%.

Table 2. Comparison of the Models based on Different Parameters

Model	mAP@0.5	mAP@0.5:0.95	Precision	Recall	F1-Score	FPS	Model Size (MB)
YOLOv8	85.90%	64.30%	86.50%	83.20%	84.80%	98	60
YOLOv9	87.20%	65.80%	87.60%	84.90%	86.20%	92	67
YOLOv10	88.40%	68.10%	89.30%	87.10%	88.20%	85	72
YOLOv10-MsA	93.10%	72.60%	92.30%	91.70%	91.90%	79	81

In addition to mAP, improvements were also observed in other critical detection metrics. YOLOv10-MsA achieved a precision of 92.3% and a recall of 91.7%, indicating a strong balance between minimizing false positives and maximizing correct detections. The combined effect of these improvements is reflected in the F1-score of 91.9%, which represents an enhancement of more than four percentage points compared with the nearest baseline.

The significant gains in accuracy can be attributed to the incorporation of the Manhattan Self-Attention (MsA) mechanism. By refining spatial feature representation, MsA allows the model to better capture subtle variations and contextual cues. This capability is especially valuable for identifying small-scale or partially occluded defects, where conventional architectures often struggle.

Overall, the results confirm that YOLOv10-MsA not only outperforms its predecessors in terms of accuracy but also offers a more robust and reliable framework for real-world defect detection tasks.

4-2- Detection Performance on Detect Types

To further evaluate the effectiveness of the proposed architecture, a class-wise mAP analysis was conducted across four primary defect categories: cracks, flashovers, contamination, and chipping. The results reveal that YOLOv10-MsA consistently outperformed the baseline models in each defect class, underscoring its robustness and adaptability in handling diverse defect patterns.

The model achieved its highest performance in crack detection, registering an impressive mAP of 94.7%, followed closely by flashover detection at 92.9%. Both defect types generally exhibit strong visual contrast in unmanned aerial vehicle (UAV) imagery, allowing the Manhattan Self-Attention (MsA) module to effectively highlight their prominent edge and contour features. This suggests that the attention-enhanced spatial sensitivity is particularly well-suited for capturing high-contrast, structurally distinct anomalies.

More notably, the improvements of YOLOv10-MsA were most significant in the detection of contamination and chipping, which are characteristically smaller in size and visually less distinct. For these challenging categories, the model attained mAP scores of 90.8% (contamination) and 91.4% (chipping). In contrast, competing models consistently recorded values below 85%. The substantial gains in these difficult-to-detect classes indicate that MsA effectively disentangles subtle visual cues from complex and noisy backgrounds, which is often a limitation of conventional detection frameworks.

Overall, the class-wise evaluation provides further evidence that the integration of MsA not only boosts general detection accuracy but also enhances the model's capacity to detect fine-grained, low-contrast defects that are critical for reliable inspection in real-world scenarios (Figure 4).

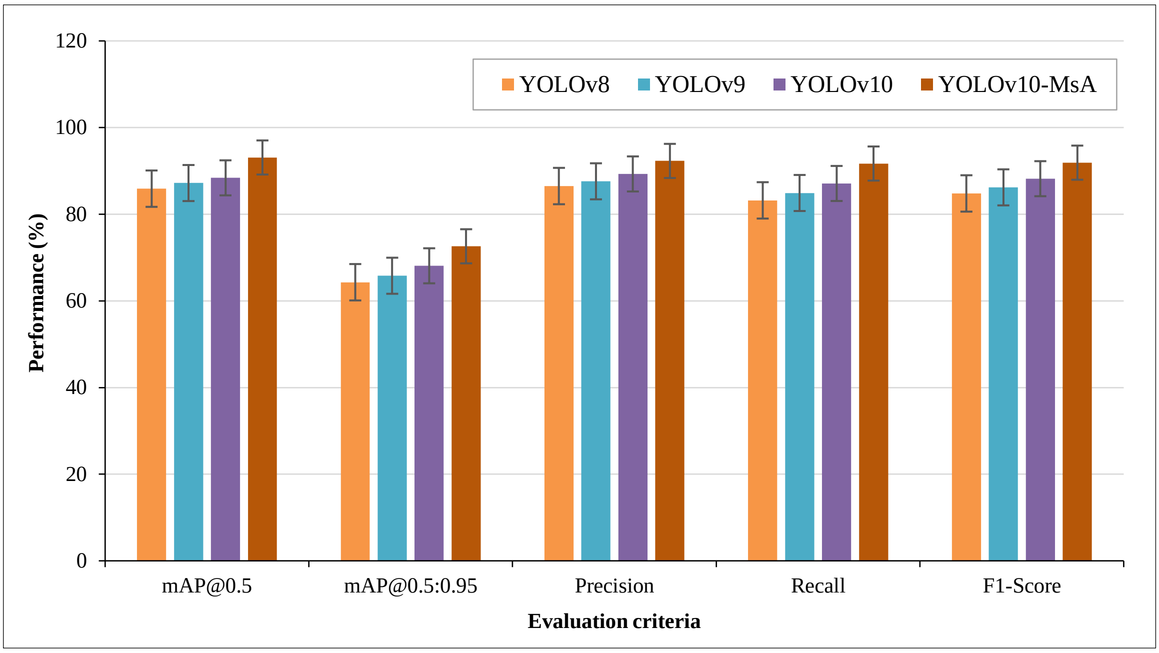

Figure 4. Performance Comparison of YOLOv8, YOLOv9, YOLOv10, and YOLOv10-MsA Models Across Key Evaluation Metrics

4-3- Qualitative Analysis and Visualization

The localization capability of YOLOv10-MsA was found to be markedly superior to the baseline models. The proposed model consistently produced tighter bounding boxes with fewer false positives, thereby offering more precise defect localization. In complex scenarios involving overlapping insulators or partially occluded views, earlier YOLO versions frequently failed to identify small defects or produced bounding boxes that were either misaligned or excessively large. By contrast, YOLOv10-MsA effectively isolated defect regions with higher spatial accuracy, further confirming the contribution of the Manhattan Self-Attention (MsA) module in enhancing multi-scale feature representation.

Another key strength of YOLOv10-MsA lies in its robustness to challenging environmental conditions. UAV imagery often suffers from backlighting, haze, fog, and variable illumination, which typically reduce contrast and hinder defect recognition. In such adverse settings, baseline models demonstrated noticeable declines in detection accuracy, often misclassifying background elements as defects. YOLOv10-MsA, however, maintained stable performance with only minimal accuracy degradation, indicating strong generalization capability.

This robustness can be attributed to the MsA's ability to emphasize critical spatial zones while suppressing irrelevant noise. For instance, non-defect background artifacts such as trees, birds, and transmission wires, which frequently distract conventional detectors, were effectively disregarded. As a result, the model delivered more reliable and context-aware predictions, ensuring accurate detection even under suboptimal imaging conditions.4.4 Inference Speed and Deployment Readiness

Despite the additional attention layers, YOLOv10-MsA retained competitive inference speed. The model achieved an average of 79 frames per second (FPS) on 640×640 input resolution during GPU-based inference. Although slightly slower than YOLOv10's 85 FPS, this performance still meets the real-time requirement for UAV-based inspections, which typically operate at 30–60 FPS.

For FPS (Frames per Second), YOLOv8 performs the fastest at 98 FPS, followed by YOLOv9 at 92 FPS. YOLOv10 is slightly slower at 85 FPS, and YOLOv10-MsA is the slowest at 79 FPS due to its more complex architecture, indicating a trade-off between speed and accuracy (see Figure 5).

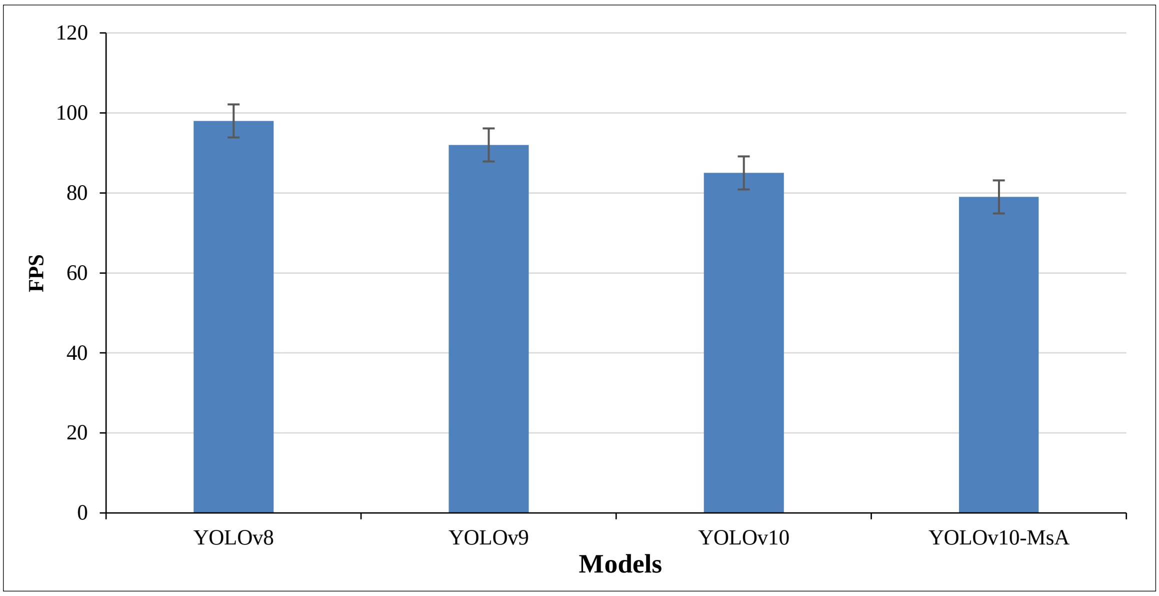

Figure 5. **Frames per Second (FPS) Comparison of YOLOv8, YOLOv9, YOLOv10, and YOLOv10-MsA Models**

The optimized YOLOv10-MsA model has a size of 81 MB, which remains sufficiently compact to allow deployment on resource-constrained platforms, such as NVIDIA Jetson Xavier devices and TensorRT-optimized drones. This compatibility highlights its practical value for real-time UAV-based inspection tasks, where both computational efficiency and detection accuracy are critical. Furthermore, techniques such as model quantization and pruning present opportunities to further reduce the model's storage and memory footprint without substantially compromising performance, thereby enhancing its suitability for embedded applications.

A comparative analysis of model sizes underscores the trade-off between architectural complexity and resource requirements. As shown in Figure 6, YOLOv8 remains the most lightweight at 60 MB, while YOLOv9 and YOLOv10 gradually increase in size to 67 MB and 72 MB, respectively. The YOLOv10-MsA variant, at 81 MB, represents the largest model in the comparison. This increase reflects the added computational overhead introduced by the Manhattan Self-Attention (MsA) mechanism, which, while contributing to a larger footprint, delivers substantial improvements in detection precision, recall, and robustness.

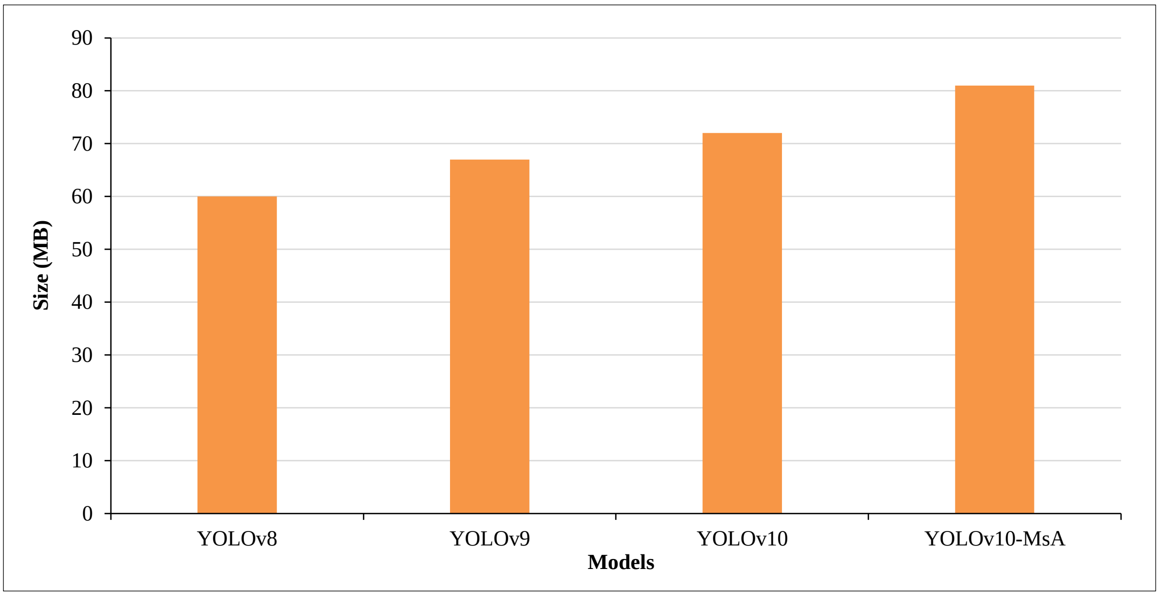

Figure 6. **Model Size Comparison (in Megabytes) of YOLOv8, YOLOv9, YOLOv10, and YOLOv10-MsA**

Overall, the results demonstrate that YOLOv10-MsA strikes a reasonable balance between model complexity and deployability, offering both high accuracy and real-world implementation feasibility.

4-4- Training Convergence and Stability

The training dynamics of YOLOv10-MsA exhibited notable improvements compared with baseline models. As illustrated by the loss curves, the proposed model achieved faster and smoother convergence than both YOLOv10 and YOLOv9. Within approximately 120 epochs, YOLOv10-MsA stabilized with only minor oscillations, indicating efficient optimization and stable learning behavior. Importantly, the validation loss plateaued early while maintaining a consistent separation from the training loss, a pattern that signifies strong generalization capability and minimal overfitting across the dataset.

Further analysis of prediction behavior revealed that the attention-enhanced architecture consistently produced higher confidence scores, reflecting greater certainty in its detections. Additionally, the model demonstrated lower variance in batch-to-batch gradient updates, a characteristic that improves training reliability. These advantages can be attributed to the Manhattan Self-Attention (MsA) mechanism, which directs the network to focus on the most informative spatial features. By guiding feature extraction more deterministically, MsA contributes to stabilized gradient flow and accelerated convergence, ultimately enabling the model to achieve higher detection accuracy with enhanced training efficiency.

4-5- Error Analysis

An error analysis was carried out to better understand the limitations of the detection models, focusing on both false positives and false negatives. The majority of false positives were linked to glare artifacts or structural components of electrical hardware that visually resemble insulators, such as spacers and clamps. These misclassifications were particularly common in YOLOv8 and YOLOv9, which often struggled to differentiate between functionally distinct but visually similar objects. In contrast, the proposed YOLOv10-MsA markedly reduced such errors, owing to its ability to perform spatial refinement through the Manhattan Self-Attention (MsA) mechanism. This improvement underscores the importance of attention-based feature enhancement in reducing contextual confusion.

Instances of false negatives were comparatively rare but did arise under particularly challenging conditions. Specifically, defects located in highly occluded regions or at the extreme edges of images were occasionally missed. Such cases are a recognized limitation across most object detection frameworks, as occlusions and edge truncations inherently reduce the visibility of critical features. Nevertheless, these errors occurred less frequently in YOLOv10-MsA compared with the baseline models. A practical solution to further mitigate this issue would involve expanding the training dataset to incorporate more examples of rare and difficult scenarios, thereby strengthening the model's resilience in edge cases.

5- Discussion

This study demonstrates how adding Manhattan Self-Attention (MsA) modules to YOLOv10 improves the detection abilities for aerial imagery of insulator defects. This segment analyzes the research findings to show their relation with previous work while evaluating both practical applications and model limitations and contributions.

5-1- Performance Gains Over Baseline and Existing Models

The proposed YOLOv10-MsA model demonstrated significant detection accuracy improvements, achieving a mAP@0.5 of 93.1% and an F1-score of 91.9%, thereby outperforming the widely used YOLOv10, YOLOv9, and YOLOv8 architectures. These results align with previous research indicating that the incorporation of attention mechanisms can substantially enhance object detector performance [6, 35]. For example, Hu et al. (2024) reported that their DGW-YOLOv8 model achieved a 3.7% increase in mAP by integrating deformable attention modules. A comparable trend is observed in this study, where the integration of the Manhattan Self-Attention (MsA) mechanism resulted in a 4.7% improvement in mAP, reinforcing the value of attention-enhanced designs.

When compared to specialized models, YOLOv10-MsA also shows advantages in balancing accuracy with inference speed. Ouyang et al. [6] introduced the E-VarifocalNet system for lightweight power grid surveillance, which delivered strong performance for detecting small-scale insulator targets. However, its real-time deployment was constrained by a detection latency exceeding 100 ms per frame. In contrast, YOLOv10-MsA achieves 79 FPS, ensuring high responsiveness and making it far more practical for UAV-based real-time inspection tasks where both speed and accuracy are essential.

Similar advances have been reported by Li et al. [36], who enhanced YOLOv8 through the integration of hybrid attention blocks, resulting in a mAP of 90.2% on insulator datasets. While effective in improving bounding box regression, their approach lacked evaluation under diverse real-world conditions. By contrast, the present study employed UAV-collected data under varying lighting conditions, different weather scenarios, and multiple levels of occlusion. This comprehensive testing strategy enhances the generalization ability and practical applicability of YOLOv10-MsA, ensuring reliable deployment in challenging inspection environments.

5-2- Comparative Detection of Defect Types

With the integration of the YOLOv10-MsA model, the detection of small and faint defects, such as contamination and chipping, achieved substantially higher accuracy than the baseline YOLO models. Earlier research [20] reported that YOLOv3 struggled to identify insulator surface contamination, primarily due to background noise and scale imbalance issues. In contrast, the Manhattan Self-Attention (MsA) mechanism employed in YOLOv10-MsA effectively isolates fine-grained spatial features, generating precise bounding boxes even for subtle and low-contrast anomalies.

Further validation of the attention-based approach can be drawn from [1], who applied a biformer-enhanced YOLOv8 for fault detection. While their model performed well in detecting flashovers and cracks, it produced a high rate of false negatives for chipped insulators, reflecting the difficulty of recognizing small-scale defects. In contrast, YOLOv10-MsA demonstrated strong and consistent performance across all defect classes, achieving mAP scores above 90% even for the most challenging categories.

These findings align with the conclusions in Apak & Farsadi [12], whose work on solar defect classification using spatial pyramid graph CNNs confirmed that spatially aware attention mechanisms significantly improve the recognition of small and intricate objects. Taken together, the results highlight that MsA not only enhances the accuracy of defect detection but also provides robust generalization across different defect types and imaging conditions.

5-3- Attention Intergration: MsA vs. Traditional Attention

This investigation validated the effectiveness of employing Manhattan Self-Attention (MsA) in place of conventional dot-product attention and Transformer-based modules. While prior studies using Transformer-enhanced architectures, such as YOLO-NAS [30], achieved competitive detection accuracy, they were often constrained by substantial memory consumption and slower inference times. By contrast, YOLOv10-MsA capitalizes on the efficiency of Manhattan distance calculations, which reduces computational overhead while still preserving a global spatial awareness critical for defect detection tasks.

These findings align with the work in Feng et al. [16], who showed that distance-based attention mechanisms in template-matching detectors delivered strong performance in few-shot learning scenarios. In a similar vein, the exponentially decaying focus on spatial proximity within MsA promotes efficient feature prioritization. This property enhances generalization capability without introducing excessive computational burden, ensuring that the model remains lightweight and practical. Importantly, this design makes YOLOv10-MsA particularly suitable for deployment on UAVs and embedded devices, where memory constraints and real-time processing requirements are significant considerations.

5-4- Practical Feasibility and Deployment

The inference speed and compact size (81 MB) of YOLOv10-MsA make it highly suitable for real-time deployment, aligning with recent trends toward lightweight detection models [5, 48]. For instance, Zan et al. [48] introduced a MobileViT-YOLO hybrid designed for low-resource platforms, which effectively reduced computational demands. However, their approach demonstrated performance trade-offs when handling high-resolution imagery, limiting its applicability in industrial inspection tasks. In contrast, YOLOv10-MsA, although slightly larger in size, sustained real-time inference at 79 FPS while also delivering superior accuracy. This balance between efficiency and precision positions the model as a robust option for industrial-grade surveillance and UAV-based inspection systems, where both reliability and responsiveness are critical.

The integration of Manhattan Self-Attention (MsA) within YOLOv10 also highlights the broader implications of this design for real-time visual recognition tasks beyond power system inspections. For example, Wang et al. [50] applied YOLOv8 to the detection of battery defects but encountered challenges in accurately localizing small electrode anomalies. The methods explored in this research, particularly the use of localized spatial attention to refine feature extraction, could be adapted to such domains, potentially improving fine-grained object detection in other high-stakes industrial applications. This demonstrates the transferability of MsA-enhanced architectures, suggesting opportunities for advancing visual recognition across multiple fields where precision and efficiency must be jointly optimized.

5-5- Limitations and Opportunities

While the proposed model delivers strong results, several limitations remain. A key challenge is the occurrence of false negatives, particularly in cases where defects are heavily occluded or located near image boundaries. Such limitations have also been reported in earlier studies in Liu et al. [21] and Zhai et al. [43] indicated that this issue is not unique to YOLO-based architectures but reflects broader challenges in defect detection. Expanding dataset diversity could help mitigate this problem. In particular, the use of synthetic data augmentation or GAN-generated samples of rare defects may enrich training distributions and enhance the model’s ability to generalize across complex scenarios.

Another limitation stems from the reliance on bounding box annotations. While efficient, this labeling approach constrains the model’s capacity to assess defect severity and fine structural variations. As argued in Finlayson & Erjavec

[26], adopting fine-grained annotation strategies, such as polygonal labeling or pixel-level segmentation masks, provides richer spatial detail that could strengthen feature learning. This transition would enable more precise quantification of defect characteristics, which is critical in high-stakes monitoring tasks.

Finally, although the integration of MsA improved detection accuracy, it also introduced additional computational complexity. For deployment on ultra-low-power UAVs or embedded edge devices, strategies such as quantization [46] and pruning [51] will likely be necessary to reduce memory footprint and inference latency. These compression techniques could ensure that YOLOv10-MsA maintains both practical efficiency and portability without sacrificing detection reliability.

6- Conclusion

Real-time power monitoring has heightened the need for automated systems to detect faults in high-voltage insulators. This study contributes to the field by introducing YOLOv10-MsA, which integrates Manhattan Self-Attention (MsA) into YOLOv10 to enhance UAV-based insulator defect detection.

The model addresses limitations of CNN-based detectors and prior YOLO versions by embedding attention modules within the backbone and neck. This design improves spatial sensitivity for detecting small or obscured features often missed in aerial inspections due to noise, altitude variation, and environmental disruption. By refining feature representation through Manhattan distance–based attention, YOLOv10-MsA strengthens defect localization without compromising inference speed.

Experimental results show YOLOv10-MsA outperforms YOLOv8 and YOLOv9 in mAP, precision, recall, and F1-score, while sustaining 79 FPS and a compact 81 MB size. These characteristics make it suitable for UAV-based inspections and edge computing. Moreover, the findings support prior evidence that MsA-enhanced CNNs can achieve performance levels comparable to transformer-based or multi-branch detectors while maintaining efficiency [1, 6, 35].

The model also demonstrates robustness across diverse defect types and environmental conditions, positioning it as a strong candidate for autonomous inspection systems. Nonetheless, challenges remain, including occasional misclassifications under severe occlusion and reliance on box-based annotations, which cannot fully capture defect complexity. Certain potential biases exist in the datasets; UAV images might not fully represent all insulator conditions, geographic and seasonal variations limited generalizability, and reliance on bounding boxes could oversimplify defect severity and morphology.

Future research should focus on three areas: (1) adopting polygonal or segmentation-based annotations for more precise labeling, (2) integrating thermal or infrared sensors to improve detection under unfavorable conditions, and (3) applying model compression techniques such as quantization, pruning, and distillation to enable deployment on ultra-low-power devices. Overall, the integration of MsA with high-performance detectors advances smart infrastructure monitoring, offering a practical pathway for deep learning applications in real-time power system inspection.

7- Declarations

7-1- Author Contributions

Conceptualization, J.Y. and N.A.M.I.; methodology, J.Y. and N.A.M.I.; validation, N.A.M.I.; formal analysis, J.Y.; data curation, J.Y.; writing—original draft preparation, J.Y.; writing—review and editing, N.A.M.I.; visualization, J.Y.; supervision, N.A.M.I.; project administration, N.A.M.I. All authors have read and agreed to the published version of the manuscript.

7-2- Data Availability Statement

The data presented in this study are available in the article.

7-3- Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

7-4- Acknowledgements

The authors would like to express their sincere gratitude to the School of Electrical and Electronic Engineering, Engineering Campus, Universiti Sains Malaysia, and all the academic and administrative staffs involved for their technical assistance throughout the research project.

7-5- Institutional Review Board Statement

Not applicable.

7-6- Informed Consent Statement

Not applicable.

7-7- Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

8- References

[1] Zhang, L., Li, B., Cui, Y., Lai, Y., & Gao, J. (2024). Research on improved YOLOv8 algorithm for insulator defect detection. Journal of Real-Time Image Processing, 21(1), 22. doi:10.1007/s11554-023-01401-9.

[2] Zheng, J., Wu, H., Zhang, H., Wang, Z., & Xu, W. (2022). Insulator-Defect Detection Algorithm Based on Improved YOLOv7. Sensors, 22(22), 8801. doi:10.3390/s22228801.

[3] Fahim, F., & Hasan, M. S. (2024). Enhancing the reliability of power grids: A YOLO based approach for insulator defect detection. E-Prime - Advances in Electrical Engineering, Electronics and Energy, 9, 100663. doi:10.1016/j.prime.2024.100663.

[4] Zhang, Y., Wu, Z., Wang, X., Fu, W., Ma, J., & Wang, G. (2023). Improved YOLOv8 Insulator Fault Detection Algorithm Based on BiFormer. 2023 IEEE 5th International Conference on Power, Intelligent Computing and Systems, ICPICS 2023, 962–965. doi:10.1109/ICPICS58376.2023.10235397.

[5] Liu, C., Yi, W., Liu, M., Wang, Y., Hu, S., & Wu, M. (2023). A Lightweight Network Based on Improved YOLOv5s for Insulator Defect Detection. Electronics (Switzerland), 12(20), 4292. doi:10.3390/electronics12204292.

[6] Ouyang, C., Zhang, H., Mu, X., Wu, Z., & Dai, W. (2023). E-VarifocalNet: A Lightweight Model to Detect Insulators and Their Defects under Power Grid Surveillance. 2023 IEEE 21st International Conference on Industrial Informatics (INDIN), 1–6. doi:10.1109/indin51400.2023.10217966.

[7] Liu, C., Wu, Y., Liu, J., Sun, Z., & Xu, H. (2021). Insulator faults detection in aerial images from high-voltage transmission lines based on deep learning model. Applied Sciences (Switzerland), 11(10), 4647. doi:10.3390/app11104647.

[8] He, M., Qin, L., Deng, X., & Liu, K. (2024). MFI-YOLO: Multi-Fault Insulator Detection Based on an Improved YOLOv8. IEEE Transactions on Power Delivery, 39(1), 168–179. doi:10.1109/TPWRD.2023.3328178.

[9] Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv Preprint, arXiv:1804.02767. doi:10.48550/arXiv.1804.02767.

[10] Gallagher, J. E., & Oughton, E. J. (2025). Surveying You Only Look Once (YOLO) Multispectral Object Detection Advancements, Applications, and Challenges. IEEE Access, 13, 7366–7395. doi:10.1109/ACCESS.2025.3526458.

[11] Hussain, M. (2024). Yolov5, Yolov8 and Yolov10: The go-to detectors for real-time vision. arXiv Preprint, arXiv:2407.02988. doi:10.48550/arXiv.2407.02988.

[12] Apak, S., & Farsadi, M. (2025). Multi-branch spatial pyramid dynamic graph convolutional neural networks for solar defect detection. Computers and Electrical Engineering, 121, 109872. doi:10.1016/j.compeleceng.2024.109872.

[13] Wu, T., Sha, T., Yao, X., Hu, J., Ma, Y., & Zhang, J. (2025). Improvement of the YOLO Series for Detecting Tower Cranes Based on High-Resolution Remote Sensing Imagery. Journal of Geovisualization and Spatial Analysis, 9(1), 8. doi:10.1007/s41651-024-00208-x.

[14] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). CBAM: Convolutional Block Attention Module. Computer Vision – ECCV 2018. ECCV 2018, Lecture Notes in Computer Science, Springer, Cham, Switzerland. doi:10.1007/978-3-030-01234-2_1.

[15] Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., & Parmar, M. (2024). A review of convolutional neural networks in computer vision. Artificial Intelligence Review, 57(4), 99. doi:10.1007/s10462-024-10721-6.

[16] Feng, C., Cao, Z., Xiao, Y., Fang, Z., & Zhou, J. T. (2023). Multi-spectral template matching based object detection in a few-shot learning manner. Information Sciences, 624, 20–36. doi:10.1016/j.ins.2022.12.067.

[17] Qi, M., Tang, R., Liu, Z., Fu, S. (2024). AYOLOv8: Improved Detector Based on YOLOv8 to Focus More on Small and Medium Objects. Advanced Intelligent Computing Technology and Applications. ICIC 2024, Lecture Notes in Computer Science, Springer, Singapore. doi:10.1007/978-981-97-5594-3_8.

[18] Adou, M. W., Xu, H., & Chen, G. (2019). Insulator Faults Detection Based on Deep Learning. 2019 IEEE 13th International Conference on Anti-Counterfeiting, Security, and Identification (ASID), 173–177. doi:10.1109/icasid.2019.8925094.

[19] Cheng, Y. (2022). Detection of Power Line Insulator Based on Enhanced YOLO Model. 2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), 626–632. doi:10.1109/ipec54454.2022.9777398.

[20] Guo, F., Hao, K., Xia, M., Zhao, L., Wang, L., & Liu, Q. (2019). Detection of Insulator Defects Based on YOLO V3.Artificial Intelligence for Communications and Networks. AICON 2019, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Springer, Cham, Switzerland. doi:10.1007/978-3-030-22971-9_25.

[21] Liu, X., Miao, X., Jiang, H., & Chen, J. (2020). Data analysis in visual power line inspection: An in-depth review of deep learning for component detection and fault diagnosis. Annual Reviews in Control, 50, 253–277. doi:10.1016/j.arcontrol.2020.09.002.

[22] Karthik, A., Kumar Sahoo, S., Kumar, A., Patel, N., Chinnaraj, P., Phaneendra Maguluri, L., Shuaib, M., & Rajaram, A. (2025). Unified approach for accurate brain tumor Multi-Classification and segmentation through fusion of advanced methodologies. Biomedical Signal Processing and Control, 100, 106872. doi:10.1016/j.bspc.2024.106872.

[23] Khanam, R., Hussain, M., Hill, R., & Allen, P. (2024). A Comprehensive Review of Convolutional Neural Networks for Defect Detection in Industrial Applications. IEEE Access, 12, 94250–94295. doi:10.1109/ACCESS.2024.3425166.

[24] Hassan, W., Hussain, G. A., Wahid, A., Safdar, M., Khalid, H. M., & Jamil, M. K. M. (2024). Optimum feature selection for classification of PD signals produced by multiple insulation defects in electric motors. Scientific Reports, 14(1), 23446. doi:10.1038/s41598-024-73196-z.

[25] Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, & Li Fei-Fei. (2009). ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition, 248–255. doi:10.1109/cvpr.2009.5206848.

[26] Finlayson, M. A., & Erjavec, T. (2017). Overview of Annotation Creation: Processes and Tools. Handbook of Linguistic Annotation. Springer, Dordrecht, Netherlands. doi:10.1007/978-94-024-0881-2_5.

[27] Dang, F., Chen, D., Lu, Y., & Li, Z. (2023). YOLOWeeds: A novel benchmark of YOLO object detectors for multi-class weed detection in cotton production systems. Computers and Electronics in Agriculture, 205, 107655. doi:10.1016/j.compag.2023.107655.

[28] Su, J., Yuan, Y., Przystupa, K., & Kochan, O. (2024). Insulator defect detection algorithm based on improved YOLOv8 for electric power. Signal, Image and Video Processing, 18(8–9), 6197–6209. doi:10.1007/s11760-024-03307-w.

[29] Arifando, R., Eto, S., Tibyani, T., & Wada, C. (2025). Improved YOLOv10 for Visually Impaired: Balancing Model Accuracy and Efficiency in the Case of Public Transportation. Informatics, 12(1), 7. doi:10.3390/informatics12010007.

[30] Terven, J., Córdova-Esparza, D. M., & Romero-González, J. A. (2023). A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. Machine Learning and Knowledge Extraction, 5(4), 1680–1716. doi:10.3390/make5040083.

[31] Panigrahy, S., & Karmakar, S. (2024). Real-Time Condition Monitoring of Transmission Line Insulators Using the YOLO Object Detection Model with a UAV. IEEE Transactions on Instrumentation and Measurement, 73, 1–9. doi:10.1109/TIM.2024.3381693.

[32] He, K., Gkioxari, G., Dollar, P., & Girshick, R., (2017). Mask, R-CNN. Computer vision (ICCV). Proceedings of the IEEE International Conference on Computer Vision (ICCV), 22-29 October, 2017, Venice, Italy.

[33] Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1137–1149. doi;10.1109/tpami.2016.2577031.

[34] Heuillet, A., Nasser, A., Arioui, H., & Tabia, H. (2024). Efficient Automation of Neural Network Design: A Survey on Differentiable Neural Architecture Search. ACM Computing Surveys, 56(11), 1–36. doi:10.1145/3665138.

[35] Hu, D., Yu, M., Wu, X., Hu, J., Sheng, Y., Jiang, Y., Huang, C., & Zheng, Y. (2024). DGW-YOLOv8: A small insulator target detection algorithm based on deformable attention backbone and WIoU loss function. IET Image Processing, 18(4), 1096–1108. doi:10.1049/ipr2.13009.

[36] Li, Z., Jiang, C., & Li, Z. (2024). An Insulator Location and Defect Detection Method Based on Improved YOLOv8. IEEE Access, 12, 106781–106792. doi:10.1109/access.2024.3436919.

[37] Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., Kamali, S., Popov, S., Malloci, M., Kolesnikov, A., Duerig, T., & Ferrari, V. (2020). The Open Images Dataset V4: Unified Image Classification, Object Detection, and Visual Relationship Detection at Scale. International Journal of Computer Vision, 128(7), 1956–1981. doi:10.1007/s11263-020-01316-z.

[38] Mumuni, F., & Mumuni, A. (2024). Segment Anything Model for automated image data annotation: empirical studies using text prompts from Grounding DINO. arXiv Preprint, arXiv:2406.19057. doi:10.48550/arXiv.2406.19057.

[39] Rizk, M., & Bayad, I. (2025). Bringing Intelligence to SAR Missions: A Comprehensive Dataset and Evaluation of YOLO for Human Detection in TIR Images. IEEE Access, 13, 17208–17235. doi:10.1109/ACCESS.2025.3529484.

[40] Lu, D., & Wang, Y. (2024). MAR-YOLOv9: A multi-dataset object detection method for agricultural fields based on YOLOv9. PLOS ONE, 19(10), e0307643. doi:10.1371/journal.pone.0307643.

[41] Liu, D. (2024). Study on insulator defect detection based on improved YOLOv8. Journal of Physics: Conference Series, 2770(1), 012009. doi:10.1088/1742-6596/2770/1/012009.

[42] Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., & Ren, D. (2020). Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. Proceedings of the AAAI Conference on Artificial Intelligence, 34(7), 12993–13000. doi:10.1609/aaai.v34i07.6999.

[43] Zhai, H., Cheng, J., & Wang, M. (2020). Rethink the IoU-based loss functions for bounding box regression. 2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), 1522–1528. doi:10.1109/itaic49862.2020.9339070.

[44] Zhang, Y. F., Ren, W., Zhang, Z., Jia, Z., Wang, L., & Tan, T. (2022). Focal and efficient IOU loss for accurate bounding box regression. Neurocomputing, 506, 146–157. doi:10.1016/j.neucom.2022.07.042.

[45] Cong, S., & Zhou, Y. (2023). A review of convolutional neural network architectures and their optimizations. Artificial Intelligence Review, 56(3), 1905–1969. doi:10.1007/s10462-022-10213-5.

[46] Saponara, S., Elhanashi, A. (2022). Impact of Image Resizing on Deep Learning Detectors for Training Time and Model Performance. Applications in Electronics Pervading Industry, Environment and Society, ApplePies 2021, Lecture Notes in Electrical Engineering, vol 866, Springer, Cham, Switzerland. doi:10.1007/978-3-030-95498-7_2.

[47] Song, Q., Zhou, Z. H., Ji, S. De, Cui, T., Yao, B. D., & Liu, Z. Q. (2024). A Multiscale Parallel Pedestrian Recognition Algorithm Based on YOLOv5. Electronics (Switzerland), 13(10), 1989. doi:10.3390/electronics13101989.

[48] Zan, W., Dong, C., Zhang, Z., Chen, X., Zhao, J., & Hao, F. (2023). Defect Identification of Power Line Insulators Based on a MobileViT-Yolo Deep Learning Algorithm. IEEJ Transactions on Electrical and Electronic Engineering, 18(8), 1271–1279. doi:10.1002/tee.23825.

[49] Souza, B. J., Stefenon, S. F., Singh, G., & Freire, R. Z. (2023). Hybrid-YOLO for classification of insulators defects in transmission lines based on UAV. International Journal of Electrical Power and Energy Systems, 148, 108982. doi:10.1016/j.ijepes.2023.108982.

[50] Wang, L., Qin, J., Yan, K., Sun, H., Sun, S., & Geng, Y. (2024). A Lightweight Insulator Defect Detection Model Based on YOLOv8. 2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT), 2349–2353. doi:10.1109/ainit61980.2024.10581815.

[51] Habib, G., & Qureshi, S. (2022). Optimization and acceleration of convolutional neural networks: A survey. Journal of King Saud University - Computer and Information Sciences, 34(7), 4244–4268. doi:10.1016/j.jksuci.2020.10.004.